Spherical Convolutions on Molecular Graphs for Protein Model Quality Assessment

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

Introduction

Processing information on 3D objects requires methods stable to rigid-body transformations, in particular rotations, of the input data. In image processing tasks, convolutional neural networks achieve this property using rotation-equivariant operations. However, contrary to images, graphs generally have irregular topology. This makes it challenging to define a rotation-equivariant convolution operation on these structures.

Method

In this work, we propose Spherical Graph Convolutional Network (S-GCN) that processes 3D models of proteins represented as molecular graphs.

Protein Graph. A protein molecule is a chain of amino acids, or residues, folded in a 3D space. We construct a graph $\mathcal{G}$ of the protein molecule by splitting the surrounding space into cells using the Voronoi tessellation method Voronota (Olechnovič & Venclovas, 2014). 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 v of the graph $\mathcal{G}$ contains a feature vector $\boldsymbol{x}_v$ associated with the corresponding protein residue. These features include one of 20 amino-acid types encoded with the one-hot representation, the solvent-accessible surface area for each residue, the volume of residue's Voronoi cell, and the "buriedness" of the residue, which is a topological distance in the graph $\mathcal{G}$ to the nearest solvent-accessible node.

Local Coordinates. The protein backbone consists of atom repetitions C, C_α, N, O. This allows us to unambiguously associate each residue with a local coordinate system. Indeed, for each residue we can define the normalized C_α–N vector as the x-axis, the unit vector lying in the C–C_α–N plane, orthogonal to x, and having positive dot product with C_α–C as the y-axis, and the vector product of x with y as the z-axis. Then, given a node v and its neighborhood $\mathcal{N}(v)$, we can associate each neighbor $u \in \mathcal{N}(v)$ 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 u onto a unit sphere in the local coordinate system of v. Now, having an unambiguous orientation for each node in the graph, we can construct a rotation-equivariant convolution operation.

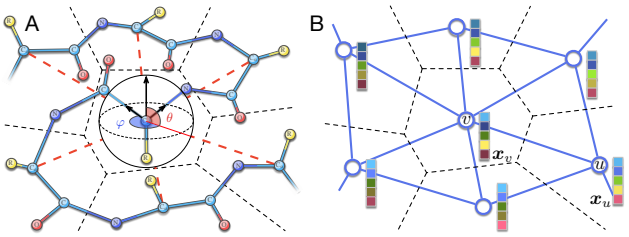

Figure 1. Schematic representation of a molecular graph. (A) 3D protein structure is partitioned into Voronoi cells, shown with the dashed lines. The central amino acid has the associated coordinate system, which is built according to the topology of its backbone (atoms C, C_α, N) with the center at the position of the C_α atom. R symbols indicate positions of the side chains. The spherical angles φ and θ of the neighboring residues are computed with respect to the local coordinate system of the central residue. (B) Graph corresponding to the Voronoi tessellation, v is the central node, u is its neighbor, $\boldsymbol{x}_v$ and $\boldsymbol{x}_u$ are the corresponding feature vectors, which are also shown with colored boxes.

Spherical Convolution. We introduce the spherical convolution operation for node v of graph $\mathcal{G}$ with some matrix function defined on a unit sphere $\boldsymbol{F} : S_1 \to \mathbb{R}^{d_1 \times d_2}$, $d_1, d_2 \in \mathbb{N}$, in the following way,

$$\boldsymbol{F} \circ v = \sum_{u \in \mathcal{N}(v)} \boldsymbol{F}(\theta_v^u, \varphi_v^u)\boldsymbol{x}_v. \tag{1}$$

Since the function $\boldsymbol{F}$ can be approximately represented as a finite sum of spherical harmonics Y_l^m and the corresponding coefficients $\boldsymbol{W}_l^m \in \mathbb{R}^{d_1 \times d_2}$ up to the maximum expansion

[*]Equal contribution [1]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France [2]Moscow Institute of Physics and Technology, 141701 Dolgoprudniy, Russia. Correspondence to: Sergei Grudinin <sergei.grudinin@univ-grenoble-alpes.fr>.

The 2021 ICML Workshop on Computational Biology.

order L,

$$\boldsymbol{F}(\theta,\varphi) \approx \hat{\boldsymbol{F}}(\theta,\varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \boldsymbol{W}_l^m Y_l^m(\theta,\varphi), \quad (2)$$

the spherical convolution can be approximated as follows,

$$\boldsymbol{F} \circ v \approx \hat{\boldsymbol{F}} \circ v = \sum_{u \in \mathcal{N}(v)} \hat{\boldsymbol{F}}(\theta_v^u, \varphi_v^u)\boldsymbol{x}_v. \quad (3)$$

Considering matrices $\boldsymbol{W}_l^m$ to be trainable parameters, we will thus learn a spherical filter.

Neural Network. Spherical Graph Convolutional Network (S-GCN) represents a sequence of spherical convolution layers combined with dropout and batch normalization layers. Spherical convoltion layer transforms the input embedding $\boldsymbol{x_v}$ of node v as follows,

$$\boldsymbol{x_v} \to \sigma\left(\hat{\boldsymbol{F}} \circ v + \boldsymbol{W}\boldsymbol{x_v} + \boldsymbol{b}\right), \quad (4)$$

where coefficients $\boldsymbol{W}_l^m$ of function $\hat{\boldsymbol{F}}$ and matrix $\boldsymbol{W}$ are trainable parameters, $\boldsymbol{b}$ is a trainable bias vector, and σ is a nonlinear activation function.

Results

Within the framework of the protein model quality assessment problem, we demonstrate that the proposed spherical convolution method significantly improves the quality of model assessment compared to the standard message-passing approach, which is a special case of the spherical convolution (4) with the maximum expansion order $L = 0$. It is also comparable to state-of-the-art 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.

The method is available at https://team.inria.fr/nano-d/software/s-gcn/.

References

Igashov, I., Pavlichenko, N., and Grudinin, S. Spherical convolutions on molecular graphs for protein model quality assessment. *Machine Learning: Science and Technology*, 2021.

Olechnovič, K. and Venclovas, Č. Voronota: A fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *Journal of computational chemistry*, 35(8):672–681, 2014.