Abstract:

RNA is one of the three major biological macromolecules that are essential for all known forms of life. The folding of RNAs into specific tertiary structures is necessary for performing specific functions and their applications in drug design.

As laboratory methods are expensive and time-consuming, it is possible to solve them by using computational methods. Due to the NP-hard of this problem, solving it requires heuristic algorithms.

There is currently only one online server for designing RNA based on its tertiary structure, which works for RNA with 100 nucleotides or less, and it takes hours to predict RNA sequences longer than 20 nucleotides.

In this thesis, a new method of deep learning is proposed to design RNA using a tertiary structure with a sequence length of more than 50. The data used in this thesis is a list of non-Redundant RNA downloaded from the NDB website[1].

In this thesis, we propose a method, named RNA3DCNN, based on 3D convolutional neural networks (3D CNN) to predict RNA sequences with lengths of more than 100 nucleotides, at a reasonable time (it takes less than 10s). In our method, the given 3D structure of RNA is broken down into each unknown nucleotide. Then, a 3D image is made using some atoms of unknown nucleotides and their neighbors based on a given RNA 3D structure. The image is fed to the model for predicting the unknown nucleotide. One by one nucleotide is generated to predict RNA sequence of 3D structure.

our model could predict the type of nucleotide for new 33 RNAs (released recently in NDB[1]) with 52% accuracy, which means much better than random accuracy (=25%).

Keywords: RNA design, Third RNA structure, Deep learning, Nucleotide sequence, Inverse folding of RNA using the tertiary structure