test

August 27, 2020

1 Introduction

1.1 Problem formulation

Given an RNA sequence of length L, we are interested in all possible secondary structures. To represent a specific secondary structure, there are three commonly used conventions: (1) undirected graph, where each node is a base in the sequence, and each edge represents base pairing. (2) upper triangular matrix (excluding the diagonal) of size $L \times L$ with binary values, where a value of 1 at (i, j) represents base pairing between sequence position i and j, and 0 represents no base paring. (3) dot-bracket notation of length L where unpaired bases are denoted as dots, and paired bases are represented as left and right brackets.

As an example, for a short RNA sequence GUUGUGAAAU, one possible structure it can take on consists of a stem and a loop, as seen in Fig **??**(a), represented by an undirected graph. Such structure can also be represented by a 10×10 upper triangular matrix with all 0's, except for positions $(1, 10), (2, 9)$ and $(3, 8)$, all being 1, as shown in Fig **??**(b). This contiguous stretch of 1's along the diagonal corresponds to the stem formed by the three base pairs: G-U, U-A and U-A. The equivalent dot-bracket notation is shown in Fig **??**(c), where the stem is represented by three pairs of left-right brackets.

1.2 Related work

2 Method

2.1 Architecture

2.2 Training

2.3 Inference

3 Result

3.1 Test set performance

3.2 Structures with pesudoknot

References

[1] Ronny Lorenz, Stephan H Bernhart, Christian Höner Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. Viennarna package 2.0. *Algorithms for molecular biology*, 6(1):26, 2011.

[2] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13):3406–3415, 2003.

[3] Mirela Andronescu, Anne Condon, Holger H Hoos, David H Mathews, and Kevin P Murphy. Efficient parameter estimation for rna secondary structure prediction. *Bioinformatics*, 23(13):i19–i28, 2007.

[4] Stefanie A Mortimer, Mary Anne Kidwell, and Jennifer A Doudna. Insights into rna structure and function from genome-wide studies. *Nature reviews Genetics*, 15(7):469, 2014.

[5] Philip C Bevilacqua, Laura E Ritchey, Zhao Su, and Sarah M Assmann. Genome-wide analysis of rna secondary structure. *Annual review of genetics*, 50:235–266, 2016.

[6] Peter Kerpedjiev, Stefan Hammer, and Ivo L Hofacker. Forna (force-directed rna): simple and effective online rna secondary structure diagrams. *Bioinformatics*, 31(20):3377–3379, 2015.

[7] Michelle J Wu. Convolutional models of rna energetics. *bioRxiv*, page 470740, 2018.

[8] Devin Willmott, David Murrugarra, and Qiang Ye. State inference of rna secondary structures with deep recurrent neural networks.

[9] Hao Zhang, Chunhe Zhang, Zhi Li, Cong Li, Xu Wei, Borui Zhang, and Yuanning Liu. A new method of rna secondary structure prediction based

on convolutional neural network and dynamic programming. *Frontiers in genetics*, 10, 2019.

[10] Linyu Wang, Yuanning Liu, Xiaodan Zhong, Haiming Liu, Chao Lu, Cong Li, and Hao Zhang. Dmfold: A novel method to predict rna secondary structure with pseudoknots based on deep learning and improved base pair maximization principle. *Frontiers in genetics*, 10:143, 2019.

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] David W Staple and Samuel E Butcher. Pseudoknots: Rna structures with diverse functions. *PLoS biology*, 3(6):e213, 2005.