Computational Methods for RNA Secondary Structure Prediction

Harrison LaBollita[1] and Petr Šulc[2]

[1]Department of Physics, Arizona State University, Tempe, AZ 85281 USA
[2]Center for Biological Physics, Arizona State University, Tempe, AZ 85281 USA

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

1 Introduction
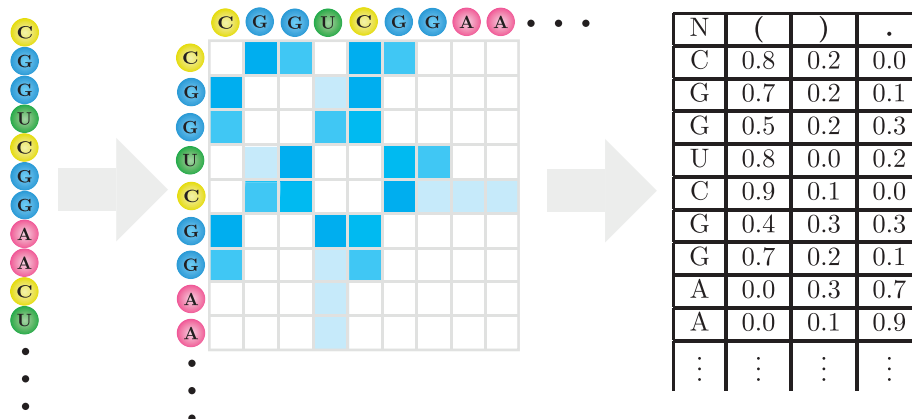
Machine learning exploratory work goes here.

In this report, we provide an implementation of the Gillespie algorithmn used to model the kinetic folding of RNA secondary structures. We have followed the work [1, 2] to realize this implementation presented below. However, in this work we introduce a slight modification. Rather than working on the base-pair level, we work at the stem level calculating the reation rate of a

stem to form, then create that stem or break depending on where we are at in the algorithmn. We have choosen to work at the stem level for computational speed purposes. We desire to combine multiple computational tools together for predicting the structure of sequences much longer than 100 nucleotides. Therefore, for our alogrithmns to be feasible we must work at the stem level.

It is extremely advantageous to combine both the Gillespie algorithm with the work of Kimichi et. al., because their work accounts for pseudoknot secondary structures, which we know (CITE) many physical RNA structures contain pseudoknots. Thus, making it vital that our algorithms do note make the canoncial (Nussinov, etc.) approximation that the structure does not contain pseudoknot stems.

## 2 Convolutional Neural Network Predicts Secondary Structure

2.1 Data and Methods

N	(	)	.
C	0.8	0.2	0.0
G	0.7	0.2	0.1
G	0.5	0.2	0.3
U	0.8	0.0	0.2
C	0.9	0.1	0.0
G	0.4	0.3	0.3
G	0.7	0.2	0.1
A	0.0	0.3	0.7
A	0.0	0.1	0.9
⋮	⋮	⋮	⋮

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

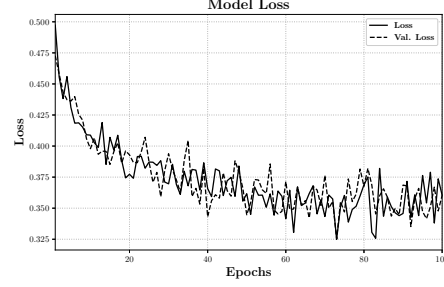### 2.2 Convolutional Neural Network Model

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.
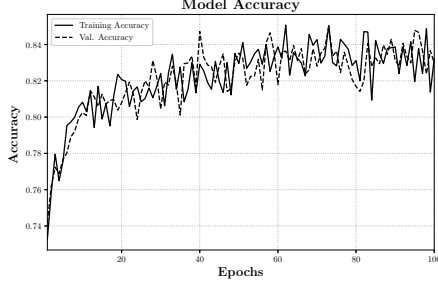
(a) Caption goes here

(b) Caption goes here

2.3 Results

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

3 Stem Level Gillespie Algorithm

3.1 Introduction to Gillespie Algorithm
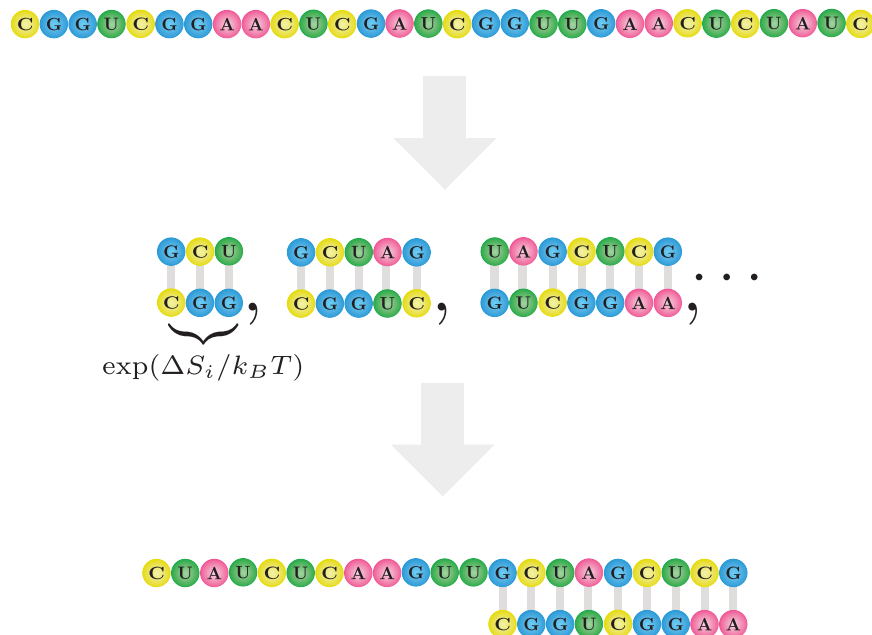
In [1], the following alogrithm is presented:

1. Extract the total flux Φ from the partial sum table.

2. Choose two random numbers r_1 and r_2 on the interval $[0, 1]$.

3. Increment the time by $\tau = -\ln(r_2)/\Phi$.

4. Identify the first *stem* which satisfies the inequality

$$\sum_{\ell=1} \phi_\ell \geq r_1 \Phi$$

   by using our table of reaction rates for each stem. We then recalculate the remainder $\Phi = r_1\Phi - \sum_{\ell=1} \phi_\ell$ on the fly.

5. Once this condition is met we choose this stem to form. We make sure it is compatible with all of the other stems in the current structure. If it is not compatible with the current structure, we remove the incompatible pieces and add our new stem. However, if adding this move requires breaking almost all of the current stems then we do not allow this move to occur.

6. Go to Step 1 until the the arbitraty cutoff time has been reached.

3

3.2 Stem Level Implementation

C G G U C G G A A C U C G A U C G G U U G A A C U C U A U C

$$\underbrace{\begin{array}{c} \text{G C U} \\ \text{C G G} \end{array}}_{\exp(\Delta S_i / k_B T)},\ \begin{array}{c} \text{G C U A G} \\ \text{C G G U C} \end{array},\ \begin{array}{c} \text{U A G C U C G} \\ \text{G U C G G A A} \end{array},\ \ldots$$

C U A U C U C A A G U U G C U A G C U C G
                    C G G U C G G A A

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

3.3 Results

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

4 Conclusion and Outlook

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie

nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

References

[1] Eric C. Dykeman. "An implementation of the Gillespie algorithm for RNA kinetics with logarithmic time update". In: *Nucleic Acids Research* 43.12 (May 2015), pp. 5708–5715. ISSN: 0305-1048. DOI: 10.1093/nar/gkv480. eprint: http://oup.prod.sis.lan/nar/article-pdf/43/12/5708/16659220/gkv480.pdf. URL: https://doi.org/10.1093/nar/gkv480.

[2] Ofer Kimchi et al. "RNA structure prediction including pseudoknots through direct enumeration of states". In: *bioRxiv* (2018). DOI: 10.1101/338921. eprint: https://www.biorxiv.org/content/early/2018/06/04/338921.full.pdf. URL: https://www.biorxiv.org/content/early/2018/06/04/338921.

[3] Hao Zhang et al. "A New Method of RNA Secondary Structure Prediction Based on Convolutional Neural Network and Dynamic Programming". In: *Frontiers in Genetics* 10 (2019), p. 467. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00467. URL: https://www.frontiersin.org/article/10.3389/fgene.2019.00467.