

Accepted Manuscript

Title: A semi-supervised learning approach for RNA secondary structure prediction

Author: Haruka Yonemoto Kiyoshi Asai Michiaki Hamada

PII: S1476-9271(15)00019-5

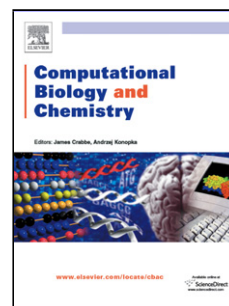
DOI: <http://dx.doi.org/doi:10.1016/j.compbiolchem.2015.02.002>

Reference: CBAC 6394

To appear in: *Computational Biology and Chemistry*

Received date: 2-1-2015

Accepted date: 3-2-2015



Please cite this article as: Haruka Yonemoto, Kiyoshi Asai, Michiaki Hamada, A semi-supervised learning approach for RNA secondary structure prediction, *Computational Biology and Chemistry* (2015), <http://dx.doi.org/10.1016/j.compbiolchem.2015.02.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A semi-supervised learning approach for RNA secondary structure prediction

Haruka Yonemoto^a, Kiyoshi Asai^{a,b}, Michiaki Hamada^{c,b,*}

^a*Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8562, Japan.*

^b*Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7, Aomi, Koto-ku, Tokyo 135-0064, Japan.*

^c*Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, 55N-06-10, 3-4-1, Okubo Shinjuku-ku, Tokyo 169-8555, Japan.*

Abstract

RNA secondary structure prediction is a key technology in RNA bioinformatics. Most algorithms for RNA secondary structure prediction use probabilistic models, in which the model parameters are trained with reliable RNA secondary structures. Because of the difficulty of determining RNA secondary structures by experimental procedures, such as NMR or X-ray crystal structural analyses, there are still many RNA sequences that could be useful for training whose secondary structures have not been experimentally determined.

In this paper, we introduce a novel semi-supervised learning approach for training parameters in a probabilistic model of RNA secondary structures in which we employ not only RNA sequences with annotated secondary structures but also ones with unknown secondary structures. Our model is based on a hybrid of generative (stochastic context-free grammars) and discriminative models (conditional random fields) that has been successfully applied to natural language processing. Computational experiments indicate that the accuracy of secondary structure prediction is improved by incorporating RNA sequences with unknown secondary structures into training. To our knowledge, this is the first study of a semi-supervised learning approach for RNA secondary structure

*Corresponding author

Email address: mhamada@waseda.jp (Michiaki Hamada)

Download English Version:

<https://daneshyari.com/en/article/6487106>

Download Persian Version:

<https://daneshyari.com/article/6487106>

[Daneshyari.com](https://daneshyari.com)