ncRNA classification using complex networks topological measures

Programa de Pós Graduação Associado em Bioinformática UFPR/UTFPR-CP

Matheus Henrique Pimenta-Zanon and Fabrício Martins Lopes

2021

Universidade Federal do Paraná and Universidade Federal do Paraná - Cornélio Procópio

Table of contents

Introduction

- RNA sequences can play various functional roles in the organism;

- RNA sequences can play various functional roles in the organism;
- In general, they can be classified as coding (mRNA) and non-coding (ncRNA);

- RNA sequences can play various functional roles in the organism;
- In general, they can be classified as coding (mRNA) and non-coding (ncRNA);
  - The mRNAs are translated in proteins necessary for cellular functions;

- RNA sequences can play various functional roles in the organism;
- In general, they can be classified as coding (mRNA) and non-coding (ncRNA);
  - The mRNAs are translated in proteins necessary for cellular functions;
  - The ncRNA have several families and their role are still not fully understood;

The role of some ncRNA is the over participation in complex human diseases, such as cancer, Alzheimer's disease, and cardiovascular diseases [1, 4].

The Next-generation sequencing (NGS), in particular the RNA-seq method, generates reads and assembling transcripts of diverse unclassified RNA.

- Propose an extension of the feature extraction method, named BASiNET [3];

Objectives

- Propose an extension of the feature extraction method, named BASiNET [3];
- Fewer thresholds are used to order to classify different RNA classes, considering four small non-coding RNA (small ncRNA) families (tRNA, IRES, miRNA, and 5SrRNA).

Materials and Methods

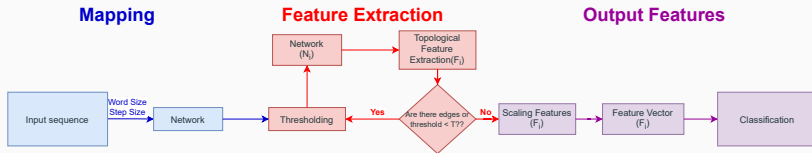The Rfam9 dataset [2] was used to provide the ncRNA sequences input to the proposed method.
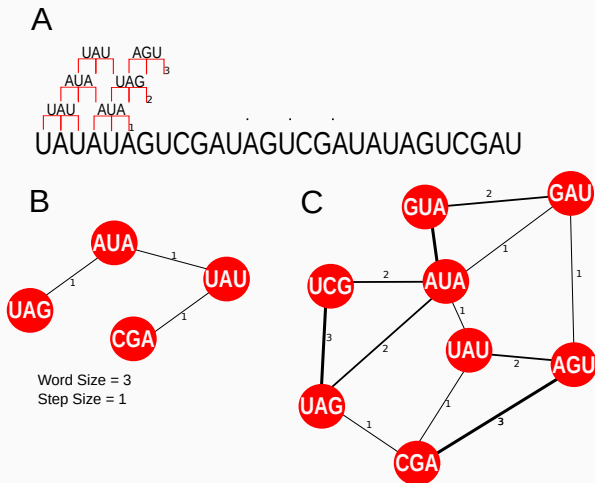
Figure 1: Overview of the BASiNET method.

Figure 2: Mapping step of the BASiNET method.

Preliminary results

Preliminary results

Using fewer thresholds values the accuracy maintains suitable results, using the Decision Tree classifier to reach 88% accuracy in a small dataset.

Preliminary results

Using fewer thresholds values the accuracy maintains suitable results, using the Decision Tree classifier to reach 88% accuracy in a small dataset.
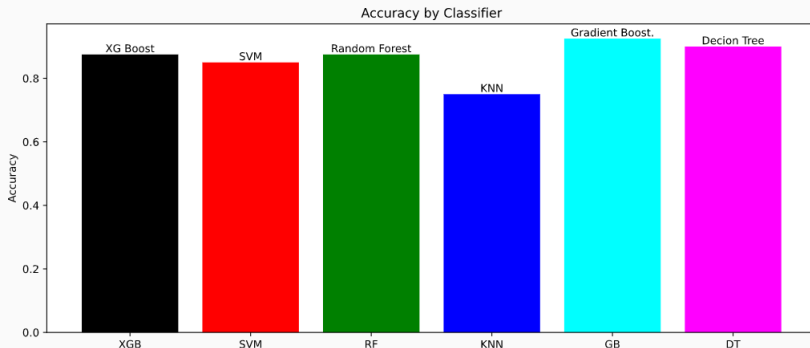

Figure 3: The highest average accuracy of the 10-fold cross validation achieved from the adopted threshold and classifiers when considering the features extracted by BASiNET.

Conclusion

Conclusion

Using fewer thresholds values the accuracy maintains suitable results;

Conclusion

Using fewer thresholds values the accuracy maintains suitable results;

The capacity to classify different ncRNA sequences, which is not shown in the first BASiNET version.

Thank you!

Thank you!

M. Esteller.
Non-coding RNAs in human disease.
Nature Reviews Genetics, 12(12):861–874, Nov. 2011.

P. P. Gardner, J. Daub, J. Tate, B. L. Moore, I. H. Osuch,
S. Griffiths-Jones, R. D. Finn, E. P. Nawrocki, D. L. Kolbe, S. R.
Eddy, and A. Bateman.
Rfam: Wikipedia, clans and the "decimal" release.
Nucleic Acids Research, 39(Database):D141–D145, Nov. 2010.

E. A. Ito, I. Katahira, F. F. Vicente, L. P. Pereira, and F. M.
Lopes.
BASiNETBiologicAl Sequences NETwork: a case study on coding
and non-coding RNAs identification.
Nucleic Acids Research, 46(16):e96–e96, Sept. 2018.

X. Shi, M. Sun, H. Liu, Y. Yao, and Y. Song.
Long non-coding RNAs: A new frontier in the study of human diseases.
Cancer Letters, 339(2):159–166, Oct. 2013.