Alignment tree for RNA pseudoknots

Quadrini M. , Merelli E. , Tesei L.

School of Science and Technology, Computer Science Division, University of Camerino, Via Madonna delle Carceri 9, 62032 Camerino; e-mail: {name.surname}@unicam.it

Ribonucleic acid (RNA) is a linear polymer of nucleotides arranged in a sequence referred to as a backbone. This sequence is made of four different types of nucleotides, known as Adenine (A), Guanine (G), Cytosine (C) and Uracil (U), and folds back on itself creating complex shapes, known as *secondary structures*. RNA secondary structures comparison is a fundamental task in several studies, among which RNA structure prediction and evolution. The comparison can currently be done efficient only for pseudoknot-free structures due to their inherent tree representation.

In this work, we introduce an algebraic language to represent both pseudoknot-free and pseudoknotted motifs that induces RNA trees permitting an efficient comparison of RNA secondary structures of any kind. For each structure, a unique *extended RNA tree* is derived from a tree grammar based on operators *concatenation*, *nesting* and *crossing*. From an extended RNA tree, an abstraction is defined in which the primary structure is neglected. The resulting structural RNA tree allows us to defined a measure of similarity calculated exploiting classical tree alignment algorithms.

References:

1) Giegerich, R., Steen, P., *Implementing algebraic dynamic programming in the functional and the imperative programming paradigm. In: International Conference on Mathematics of Program Construction, Springer,* **2002***,* 1-20

2) Bille, P., *A Survey on Tree Edit Distance and Related Problems. Theoretical Computer Science* 337*,* **2005**, -217-23

3) Quadrini, M., Culmone, R., Merelli, E.: *Topological Classification of RNA Structures via Intersection Graph. In: International Conference on Theory and Practice of Natural Computing, Springer,* **2017**, 203–215