EACH | campus capital USP LESTE
Escola de Artes, Ciências e Humanidades
Universidade de São Paulo

Graduate Program in Information Systems
PPgSI

6th PPgSI's Dissertations Workshop 2019

Prediction of Three-Dimensional Structures for RNA Molecules using Game Theory

Author:	Jhonatan Silva da Silva			
Advisor:	Daniel de Angelis Cordeiro, Dr.			
Co-advisor:				
Research lines:	[X] Systems Development and Management		[ ] Systems Intelligence	
Research areas:	[ ] Database	[X] Software engineering	[ ] Artificial intelligence	[ ] Graphics processing
	[ ] Information technology management	[ ] Human-Computer Interaction	[ ] Pattern recognition	[ ] Optimization
Application areas:	[ ] Enterprise environments / Business processes	[X] Bioinformatics	[ ] Biometrics	[ ] Mobile devices
	[ ] Economy	[ ] Education / Distance learning	[ ] E-government	[ ] Internet / Social Networks
	[ ] Games / Serious games	[ ] Linguistics / Natural Language	[ ] Cheminformatics	[ ] Robotics
	[ ] Health	[ ] Other Which? _____		[ ] General*

* Check "GENERAL" if your research is not directed at any particular application area, i.e., it is exclusively focused on the research area stated above. To be sure this is your case, make sure that the result you expect to produce contributes indirectly to **any** area of application of information systems/computer science. For example, your research enhances a software testing technique, proposes a new data structure or establishes a heuristic to parameterize a machine learning algorithm. Note that in such cases, the output produced can be useful for any social problem in which automated solutions are needed. **(REMOVE THIS LINE IN THE *CAMERA READY* OF THIS DOCUMENT)**

Period in the program (at the workshop date):	[ ] 2nd semester	[ ] 3rd semester	[ ] 4th semester	[X] 5th semester
Qualifying:	[X] Qualifying held in: 23/11/2018		[ ] Plan for qualifying in: dd/mm/yyyy	
Defense:	Deadline for deposit: 27/01/2020		Plan for defending in: dd/mm/yyyy	
Publications associated with the master's project:	**No publications to date.**			

Keep this information organized on a single page. Please, remove this comment in the *camera ready* version.

The research project summary

Context:

Ribonucleic Acids (RNA) are molecules present in living beings, responsible for functions such as protein synthesis through the translation of the genetic code catalysis of biochemical reactions and performance of regulatory individual. The 3D structure of RNAs is important, especially when studying their functions.

These 3D structures are used, for example, in the production of certain drugs and can provide the discovery of disease-causing mutations. To obtain such structures with the least probability of errors, experimental techniques are used. However, they are time consuming and expensive. Computational techniques began to be developed in an attempt to predict such structures, although, there are limitations such as runtime and/or size of the molecule to be predicted.

A recent technique named GARN (Game Algorithms for RNa 3D sampling) has shown to supply these limitations using Game Theory concepts, where clusters of smaller RNA units called nucleotides are represented by players, and the game has rules for each type of player using a potential energy function based on in knowledge.

Research problem:

The technique GARN left a gap related to the prediction of small molecules, which differ greatly when compared to approved structures arising from experimental techniques. One possible solution cited by the authors is to change the game's settings for finer structures, that is, visually more approximated from structures obtained by experimental techniques

Research objective:

Improve the technique developed in the studies by Boudard et al. (2015) and Boudard et al. (2017) to predict the three-dimensional structure of RNA molecules using Game Theory, aiming to obtain a more refined three-dimensional structure, that is, with better approximation of a laboratory derived structure.

Characteristics of the proposed solution:

Two algorithms are being used in these experiments:
EXP3 (EXPonential EXPloration-EXPloitation)
UCB (Upper Confidence Bound)

Theoretical foundations:

By refining the technique developed in the studies by Boudard et al. (2015) and Boudard et al. (2017) to obtain a more refined three-dimensional structure, it is possible to keep the efficiency of the technique in terms of execution time. It is therefore expected that improvements in the technique proposed by the authors do not increase the complexity of the algorithms significantly.

Correlated works:

- BOUDARD, Mélanie et al. GARN: sampling RNA 3D structure space with game theory and knowledge-based scoring strategies. **PloS one**, v. 10, n. 8, p. e0136444, 2015.
- BOUDARD, Mélanie et al. GARN2: coarse-grained prediction of 3D structure of large RNA molecules by regret minimization. **Bioinformatics**, v. 33, n. 16, p. 2479-2486, 2017.
- BERNAUER, Julie et al. Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. Rna, v. 17, n. 6, p. 1066-1075, 2011.

Validation

Evaluation for each molecule

For each three-dimensional structure modeled by the technique, it will be verified how similar it resembles to the derived structure in using the graph. For this, a calculation will be used to measure based on the distance from the similarity of the protein structure, called RMSD (Root Mean Square Deviation). RMSD is a form of quantitative evaluation widely used to measure similarity between two atomic coordinates overlapping (KUFAREVA; ABAGYAN, 2011).

Comparison with other techniques

When performing the Literature Review, 14 studies were found.

Of this total, 8 techniques were relevant to be tested in this study, not only referenced by other studies, but also used for the purpose of comparing the results. Thus, in this study, the techniques NAST (JONIKAS et al., 2009), FARNA (DAS; BAKER, 2007), RNAComposer (POPENDA et al., 2012), iFoldRNA (KROKHOTIN; HOULIHAN; DOKHOLYAN, 2015), ERNWIN (KERPEDJIEV; SIEDERDISSEN; HOFACKER, 2015), RNAJAG (LAING et al., 2013) and 3dRNA (WANG et al., 2017) will be used for the purpose of comparison of the results.

Limitations, risks, and threats:
The database to be used in this research is limited, and this may influence in the final result, considering that this base is used to obtain the conformations of the atoms and, consequently, perform the analyzes to determine the game's settings.

Also, the tools to be used for results comparison purposes are available online, and we will use to compare the results only to the techniques widely known and already used by other studies for comparison purposes.

Scientific contribution:
The technique, when refined for a more refined approach, is expected to maintain a runtime comparable to other techniques in the literature, as in the study of techniques GARN (BOUDARD et al., 2015) and GARN2 (BOUDARD et al., 2017). Having the expected result, the study is expected to contribute to the area of bioinformatics, improving a technique for predicting three-dimensional structures with the attempt to decrease the need for derivation of the structures in the laboratory.

Technical contribution (if pertinent):

The research method			
Genre (choose ONE)	[ ] Theoretical research	[X] Practical research	[ ] Empirical research [ ] Methodological research
Nature (choose ONE)	[X] Basic research	[ ] Applied research	
Approach (choose ONE)	[X] Quantitative research	[ ] Qualitative research	[ ] Quali-quanti research
Literature review* (you can choose more than one)	[ ] Narrative review [ ] Descriptive review [ ] Scoping review	[ ] Meta-analysis [X] Qualitative systematic review [ ] Umbrella review	[ ] Theoretical review [ ] Realistic review [ ] Critical review
Main technical procedure (choose ONE)	[X] Experimental research [ ] Bibliographic research [ ] Documental research [ ] *Ex-post-facto* research	[ ] *Survey* [ ] Case study [ ] Participatory research [ ] Research-action	[ ] Ethnographic research [ ] Grounded theory [ ] Design science [ ] Other Which? _____
Data analysis (you can choose more than one)	[X] Descriptive statistics [ ] Inferential statistics	[ ] Statistical test [ ] Content analysis	[ ] Discourse analysis [ ] Others: _____

* Definition of types of literature reviews established by Paré, G., Trudel M-C., Jaana M., Kitsiou, S. Synthesizing Information systems knowledge: A typology of literature reviews. In: Information & Management 52, p. 183-199, 2015. DOI: 10.1016/j.im.2014.08.008

Next steps:
Applying different scores and strategies to helices, according to the energy potential obtained;
At the same time, comparing the new structures to the old technique (GARN) structures, trying to obtain sufficiently better results

Keep the information regarding the RESEARCH PROJECT SUMMARY and the RESEARCH METHOD organized in no more than two pages. Remove that remark in the *camera ready* version.

Optional: Provide a graphic layout that shows aspects of your research. For example: a flow-chart for building your solution or an infographic for your research proposal. If necessary, use the fourth page.