244 New Computer Science Building
Stony Brook University
Stony Brook, NY 11794
☎ (631) – 974 – 8156
✉ falmodaresit@cs.stonybrook.edu
https://www.linkedin.com/in/fataltes
https://github.com/fataltes
https://fataltes.github.io

Fatemeh Almodaresi

Research Interests

Computational Biology
Algorithms and Data Structures
Data Mining, Pattern Recognition, and Data Analysis

Education

2015-present **Ph.D**, *Computer Science Department, Stony Brook University (SBU)*, NY, USA.
Advisor – Prof. Rob Patro

2009-2011 **MS**, *School of Computer Engineering, Iran University of Science and Technology (IUST)*, Tehran, Iran.
Advisor – Prof. Jahed Motlagh

2004-2009 **BS**, *School of Electrical & Computer Engineering (ECE), University of Tehran*, Tehran, Iran.

Selected Research Projects

Aug. 2017 - Present **Mantis**, *"A fast, small, and exact large-scale sequence-search index"*, Computational Biology Lab., SBU, `https://github.com/splatlab/mantis`.
Mantis is a space and time efficient data structure to index and query large collections of raw sequencing read experiments. The index is based on colored de Bruijn graph representation and therefore supports graph traversal, bubble calling and other graph-based computational and biological analysis.

Jun. 2017 - Jan. 2018 **Pufferfish**, *"A space and time-efficient compacted de Bruijn graph index"*, Computational Biology Lab., SBU, `https://github.com/COMBINE-lab/pufferfish`.
Pufferfish is an efficient data structure for indexing colored compacted de Bruijn graphs. This tool can achieve a balance between time and space resources by making use of succinct data structures and minimum perfect hash function. Pufferfish provides the underlying data structure for mapping short sequencing reads to a huge population of references while keeping the mapping information for each reference individually. Plan to submit to Recomb 2018.

Apr. - Jun. 2017 **Rainbowfish**, *"A succinct colored de Bruijn graph data structure"*, Computational Biology Lab., SBU, `https://github.com/COMBINE-lab/rainbowfish`.
This tool provides a new data structure to store and query colored de Bruijn graphs that in case of large data sets improves storage by more than twenty times compared to state-of-the-art tools without hurting performance of the queries.

Nov. 2016 - Present **Grouper**, *An extension to "Rapid Clustering" tool*, Computational Biology Lab., SBU, `https://github.com/COMBINE-lab/grouper`.
Grouper is a tool for clustering contigs of a de novo transcriptome assembly. We improved the accuracy of clustering by making use of orphan reads, for which each end of the pair is mapped to a different reference sequence (accepted to Bioinformatics).

Aug 2016-Jan 2017 **MLDD**, *"Multi-Level Distribution Detection"*, Data Science Lab., SBU.
Using statistical tests and classification models such as NaiveBayes we show how distribution of NLP features in social media changes in different levels of analysis (county, user, and message). This can highly affect prior assumptions for further text analysis as we show that central-limit theorem could be applied in social media language analysis as well.

2013-2014 **AutismFD**, *"A game to improve face emotion detection in children with Autism"*.
Beside collaboration with psychology students to design the method, I also implemented the idea as a tool in C# language. This package was used in a treatment center to help children with Autism to identify face emotions and track their progress over time.

Publications

[1] Prashant Pandey, Fatemeh Almodaresi, Michael A Bender, Michael Ferdman, Rob Johnson, and Rob Patro. Mantis: A fast, small, and exact large-scale sequence-search index. *Cell Systems*, 2018.

[2] Laraib Malik, Fatemeh Almodaresi, and Rob Patro. Grouper: Graph-based clustering and annotation for improved de novo transcriptome analysis. *Bioinformatics*, 1:8, 2018.

[3] Fatemeh Almodaresi, Hirak Sarkar, Avi Srivastava, and Rob Patro. A space and time-efficient index for the compacted colored de bruijn graph. *Bioinformatics*, 34(13):i169–i177, 2018. (appeared in the proceedings of ISMB 2018).

[4] Fatemeh Almodaresi, Prashant Pandey, Michael Ferdman, Rob Johnson, and Rob Patro. An efficient, scalable and exact representation of high-dimensional color information enabled via de bruijn graph search. *bioRxiv*, 2018. (accepted in RECOMB 2019).

[5] Mohsen Zakeri, Avi Srivastava, Fatemeh Almodaresi, and Rob Patro. Improved data-driven likelihood factorizations for transcript abundance estimation. *Bioinformatics*, 33(14):i142–i151, 2017. (appeared in the proceedings of ISMB 2017).

[6] Fatemeh Almodaresi, Lyle Ungar, Vivek Kulkarni, Mohsen Zakeri, Salvatore Giorgi, and H Andrew Schwartz. On the distribution of lexical features at multiple levels of analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 79–84, 2017.

[7] Fatemeh Almodaresi, Prashant Pandey, and Rob Patro. Rainbowfish: A Succinct Colored de Bruijn Graph Representation. In *17th International Workshop on Algorithms in Bioinformatics (WABI 2017)*, volume 88 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:15, 2017.

Work Experiences

Jun-Aug 2016 **Member of the NLP Team**, *Third Frederick Jelinek Memorial Summer Workshop (JSALT)*, Baltimore.
JSALT is a well-known summer workshop in Language and Speech organized by JHU each year. .
During the project, we worked on analyzing and forecasting social media user's psychological state based on their language in their posts using statistical methods such as significance tests and time series models such as ARMA and ARIMA.

Jan-Aug 2015 **Senior Designer and Developer**, *Nexeven AB*.
Nexeven AB is a Swedish company and a niche player in the online video broadcasting field.

2011-2015 **Team Supervisor, Senior Designer and Developer**, *Tosan Intelligent Data Miners Co. (TIDM)*, Data Mining Development Team.
TIDM is the first solution provider for fraud detection and anti-money laundry in banking section in Iran, a Subsidiary of Tosan Company.

- **Customer Relationship Management System** [2014]
  In this project we use statistical and data mining methods to calculate customer's RFM, CLV, and churn probability.
- **Data mining Module, Operational Intelligence System** [2014]
  This module, developed in PLSQL, uses Statistical and Mining Methods such as regression models, error functions, k-means, and SVM to detect fraudulent transactions online in the stream of transactions.
- **Customer Name Similarity Detection Module** [2013]
  As a part of Anti-money Laundry System, this module uses natural language algorithms to detect accounts with similar names. The whole system is developed in PLSQL and now operational in many private banks in Iran including Eghtesad-Novin and Ansar Bank.
- **Unsupervised Fraud Detection System, Version 1 & 2** [2011-2014]
  Version 1 which is fully designed and developed by myself is now operational in Saman Bank, Ansar Bank, and Mehr-e-Eghtesad Bank in Iran. Version 2 is now installed in Eghtesad-Novin Bank.

2009-2011 **Java and UI Developer** , *Tosan Co.*
Tosan Company is a pioneer company for total banking solutions with more than fifteen Iranian financial institutes in its customer list. As a member of a team of nearly 20 people, I participated in developing the UI of Internet banking system.

Honors & Awards

2018 **ISMB2018 Conference Travel Fellowship**.

2016 **CS Department Best TA Award**.

2015 **Special CS Department Chair Fellowship**.

Teaching Experiences

Fall 2017 **Teaching Assistant**, *Computational Biology*, Stony Brook University.

Spring 2017 **Teaching Assistant**, *Machine Learning*, Stony Brook University.

2013 **Teacher**, *C++ Programming Language*, Farzanegan High School [NODET].

2013 **Teacher**, *Developing simple motion detection algorithms in MATLAB*, Farzanegan High School [NODET].

Fall 2008 **Teaching Assistant**, *Artificial intelligence*, University of Tehran.

Skills

Programming Languages **Python (expert), Java Core(expert), C++ (expert), R (familiar), MATLAB (familiar), NetLogo (familiar), C# (familiar)**.

Libraries and Frameworks **Popular Python Libraries (numpy, pandas, scipy.stats, sklearn), ShinyR, Spring Framework, Hibernate, Play Framework**.

Databases **Oracle (expert), MySQL (expert), MongoDB (familiar)**.

Other Tools **Git, Atlassian Jira, Atlassian Confluence, ThoughtWorks Go, Anaconda Platform, Pycharm, Jupyter Notebook, IntelliJ IDEA**.