Serhan Yılmaz

www.serhanyilmaz.com　　　　sxy600@case.edu　　　　github.com/serhan-yilmaz

Education

08/2018 - Present　　Ph.D. in Computer Science
Case Western Reserve University, Cleveland (Ohio, US)
Research Advisor: Mehmet Koyutürk
CGPA: 4.00/4.00

09/2016 - 07/2018　　M.S. in Computer Engineering
Bilkent University, Ankara (Turkey)
Thesis Advisor: A. Ercüment Çiçek
CGPA: 3.86/4.00

09/2013 - 06/2016　　Minor degree in Information Systems (Dept. of Computer Engineering)
Middle East Technical University, Ankara (Turkey)
CGPA: 3.50/4.00

09/2012 - 06/2016　　B.S. in Electrical and Electronics Engineering (Major)
Middle East Technical University, Ankara (Turkey)
CGPA: 3.30/4.00 (Specialization: Computer Architectures Area)

Research Experience

08/2018 - Present　　**Research Assistant, Case Western Reserve University**

Currently working on three research problems:

(1) **Kinase activity inference:** Phospho-proteomics data in combination with computational algorithms allows the inference of kinase activities, which leads to the development of kinase inhibitors for targeted therapy of various diseases like cancer and Alzheimer's disease. To improve the robustness of the inference (to missingness in input data), developed an algorithm named RoKAI that integrates various sources of functional information (in the form of a heterogeneous graph) to capture coordinated changes in signaling. To make it more accessible, developed a web application to perform RoKAI through a user-friendly online interface.

(2) **Mining electronic health records:** In collaboration with a domain expert on domestic violence, used data mining techniques to analyze the electronic health records of women who are victims of intimate partner violence (IPV) to uncover health problems that are potential consequences of IPV, particularly for vulnerable populations like pregnant or older women.

(3) **Benchmarking link prediction algorithms:** In collaboration with another graduate student, we investigate potential biases in the evaluation of link prediction algorithms in various contexts (recommendation systems, social networks, systems biology).

09/2016 - 07/2018　　**Research Assistant, Bilkent University**

- Worked on the problem of high-dimensional feature selection for genomic datasets. This is an important problem to uncover which of the $\sim 10^6$ mutations in genome are responsible for disease phenotypes.
- Developed an algorithm, SPADIS, that performs a sub-modular optimization to select a predictive SNP set for a given phenotype while using a network to reduce the search space (in order words, it performs a graph regularized regression).
- Implemented the algorithm in C++ for efficiency, and ported to MATLAB for the analysis
- Prepared a comprehensive framework for benchmarking in $\sim 10^7$ different settings (different phenotypes, parameters, cross-validation folds etc.) that is capable of running in parallel on 40 cores.
- Additional responsibilities included the maintenance of the lab server as a sysadmin.

Teaching Experience

09/2018 - 06/2019   **Teaching Assistant, Case Western Reserve University**

Fall 2018, EECS-233 2nd year data structures course, homework grader

Spring 2019, EECS-454 graduate level algorithms, grader and course coordinator

Fall 2019, DSCI-343, 3rd year data science course, prepared course materials and graded homeworks

09/2016 - 07/2018   **Teaching Assistant, Bilkent University**

Fall 2016, CS-223 2nd year digital design course, TA in laboratory experiments, project grader

Spring and Fall 2017, CS-201 2nd year fundamental data structures course, homework grader

Publications

- Hacialiefendioglu, A. M., **Yilmaz, S.**, Koyuturk, M. & Karakurt, G. (2020). Co-occurrence Patterns of Intimate Partner Violence. *Pacific Symposium on Biocomputing.*

- **Yilmaz, S.**, Alghamdi, B., Singuri, S., Hacialiefendioglu, A. M., Ozcan, T., Koyuturk, M. & Karakurt, G. (2020). Identifying health correlates of intimate partner violence against pregnant women. *Health information science and systems* (link).

- **Yilmaz, S.**, Tastan, O., and Cicek, A. E. (2018). SPADIS: An Algorithm for Selecting Predictive and Diverse SNPs in GWAS. *IEEE/ACM transactions on computational biology and bioinformatics* (link)*.*

Preprints/Manuscripts under review

- **Yilmaz, S.**, Ayati, M., Schlatzer, D., Cicek, A. E., Chance, M., & Koyuturk, M. (2020). Robust Inference of Kinase Activity using Functional Networks. Preprint in bioRxiv. (link)

- **Yilmaz, S.**, Gunay, E., Lee, D. H., Whiting, K., Silver, K., Koyuturk, M. & Karakurt, G. (2020). Mining Electronic Health Records to Identify Adverse Health Correlates of Intimate Partner Violence against Older Women.

- Hacialiefendioglu, A. M., **Yilmaz, S.**, Koyuturk, M. & Karakurt, G. (2020). Data-driven identification of subtypes of intimate partner violence.

- Ayati, M., **Yilmaz, S.**, Schlatzer, D., Maxwell, S., Li., M, Chance, M. R. & Koyuturk, M. (2020). Co-Phosphorylation Networks Capture Functional Relationships among Protein Phosphorylation Sites.

Talks and Poster Presentations

- **Robust Inference of Kinase Activity using Functional Networks.**
  ISMB 2020, Virtual Conference, on July 13, 2020. 20-min talk and poster presentation (link).
  HIBIT 2020, Virtual Conference, on October 22, 2020. 20-min talk.

- **Spadis: An algorithm for selecting predictive and diverse snps in GWAS.**
  ISMB 2018 at Chicago, Illinois, United Stated on July 8, 2018. 10-min talk (recording) and poster (link)

Academic Service

Serving as a reviewer for various journals and conferences since 2017:

- Journal of Proteome Research, ISMB/ECCB, Recomb, Recomb-CCB. ACM-BCB. CNB-MAC

Honors and Awards

- 2016, Bulent Kerim Altay Award for academic excellence, given by the Department of Electrical Engineering, Middle East Technical University

- 2015, Best Project Award in EE314 Digital Electronics Laboratory course

- 2012, Turkey University Entrance Exam (LYS) degree 642nd out of 2,137,000

Computing Skills

Programming Languages:	**C++**, **Java**, C#, C, JavaScript
Data Analysis Tools:	**MATLAB**, **R**, Python
Packages/Libraries:	R-shiny, tidyverse, ggplot2, tensorflow, numpy, pandas, matplotlib, bokeh
Other Languages:	SQL, HTML, CSS, Bash, UML, Verilog, VHDL, Assembly (Motorola 6800)
Development Tools (IDEs):	Visual Studio (C/C++, C#), Netbeans (Java), RStudio (R)
Miscellaneous:	LaTeX, R Markdown, Markdown, Office tools, Photoshop, Unity

Software Projects (Professional)

- 2020, **CrossTabulator**: Developed an R shiny web application to generate summary tables from input data easily. It is particularly designed to create demographics tables for use in scientific journals. https://crosstabulator.com/

- 2020, **NetworkVisualizer**: Built a graph visualization tool for MATLAB to create some pretty network figures (see examples). Prevented the node overlaps in graph layout by adapting the Fruchterman-Reingold algorithm for variable node sizes. Used a pipe-like structure to make the graphs easily customizable (inspired from the ggplot2 library in R).

- 2020, **ClustXplorer**: An interactive web-based tool for visualizing clusters in a hierarchical structure. It is implemented in Python with the help of bokeh library for interactive layout.

- 2020, **RoKAI**: A computational tool for robust inference of kinase activities using functional networks. MATLAB.

- 2020, **RoKAI App** : A web application for performing RoKAI through a user-friendly online interface. http://rokai.io

- 2019, **Robo-Doc** (**Hackathon Project for HackCWRU)**: Built an application for automated prediction of differential diagnosis for a pool of 156 diseases and 330 symptoms. Applied Bayesian inference with active learning to select an optimal set of questions to ask a patient. Application implemented in Java and benchmarking done in MATLAB.

- 2018, **SPADIS**: An algorithm for selecting a predictive and diverse SNP set in genome wide association studies (GWAS). Implemented in C++ and ported to MATLAB for easy access (github link).

- 2015, **Flight Simulator (IARC Competition)**: A java application to simulate the behavior of an autonomous aerial robot for the international aerial robotics competition in 2015 (for more information link).

Software Projects (Personal)

- 2018, **Dota2HeroMatchup:** A java application (link) to help select statistically winning heroes that counter enemy picks for the game Dota 2. Implemented a web crawler to obtain matchup data and estimated the team win rates based on a Bayesian model.

- 2017, **AutoGrader**: An application to automatically test and grade coding assignments in C++. It prepares a summary Excel file reporting compile errors, time-outs, memory leaks, incorrect logic and fatal errors.

- 2016, **Family Tree**: A java application (link) that provides an efficient way to store and view family trees graphically.

- 2016, **Hatman**: An indie game made in Java with 2D graphics. The aim is to survive as long as possible while running from hazards that get increasingly more difficult (see github for more information).

- 2016, **QueueSimulator**: A java application (link) that simulates the G/G/k queue systems in computer networks to estimate the latency and potential packet losses.

- 2015, **Agar Simulator**: A game/simulator (link) inspired from the game *agar.io*. Implemented in Java.

Course Projects

- 2019, **Data Science Course Project**: Applied data cleaning and data imputation techniques on an healthcare dataset and performed A/B hypothesis testing to compare two groups of clinical practices for their effectiveness. The analysis is done in R using tidyverse, dplyr, ggplot2 libraries.
- 2019, **Causal Inference Course Project**: Assessed the causal effect of obesity on diabetes using a generalized linear model (GLM) and ANOVA after adjusting for relevant confounders. The analysis is done using R.
- 2018, **Data Mining Course Project**: Implementation and performance evaluation of five classification algorithms: (i) Random forest, (ii) PCA classifier, (iii) LDA classifier, (iv) k-NN, (v) Support vector machines (SVM). MATLAB.
- 2017, **Deep Learning Course Project**: Utilizing temporal information in videos to improve the bounding-box predictions generated by still-image detectors, by employing recurrent neural networks using TensorFlow toolbox.
- 2017, **Machine Learning Course Project**: Mixture of experts model is employed with multilayer perceptron (MLP) classifiers each trained specifically to learn a particular class. It is implemented in MATLAB.
- 2016, **Pattern Recognition Course Project**: Project employs linear classification methods into a regression problem by considering cost-sensitive decisions implemented in MATLAB.
- 2016, **EEE Senior Design Project**: Development of an autonomous system that can detect and lock onto the target and shoot towards it with a bullet like object using coil gun technology.
- 2015, **Digital Signal Processing Project**: Autonomous recognition of vowel letters in MATLAB.
- 2015, **Digital Electronics Lab. Project**: Design and implementation of an arcade game on FPGA using Verilog HDL in addition to its prototype in Java.

Language Skills

- English: Advanced - TOEFL iBT 109/120
- Turkish: Mother Tongue

Hobbies and Interests

- ❖ Solving problems on Project Euler (solved 160 problems out of 728 so far)
- ❖ Participating in online code/algorithm challenges (e.g., Google kickstart, Leetcode).
- ❖ Analyzing video games mechanics to find the most optimal play
- ❖ Developing video games (especially 2D arcade games)
- ❖ Designing logos, posters or anything else that encourages team spirit
- ❖ Creating good looking figures, plots or charts