Hossein Soleimani

Contact Information

Robust Machine Intelligence and Control Lab
114 Electrical Engineering West
Pennsylvania State University
University Park, PA 16802

✉ hsoleimani@psu.edu
www https://hsoleimani.github.io
linkedin.com/in/hsoleimani
github.com/hsoleimani

Education

Pennsylvania State University, University Park, PA

Ph.D., Electrical Engineering, January 2012 - June 2016

- Minor: Statistics
- Advisor: Dr. David J. Miller

University of Tehran, Tehran, Iran

M.Sc., Electrical Engineering, 2011

Ferdowsi University of Mashhad, Mashhad, Iran

B.Sc., Electrical Engineering, 2008

Research Interests

Machine learning, Probabilistic graphical models, Approximate posterior inference, Text processing, Statistical modeling, Information retrieval, Signal processing, Bioinformatics.

Honors and Awards

- Dr. Nirmal K. Bose Dissertation Excellence Award, Department of Electrical Engineering, Pennsylvania State University, 2016
- German Academic Exchange Service (DAAD) scholarship for a 6-month research program at University of Rostock, Germany, July-December 2010.
- Ranked second in national university exam for M.Sc. program, Iran, 2008.

Programming and Software

- Proficient in C, Python, MATLAB.
- Experienced in R, Linux, LaTeX, and Simulink.
- Familiar with Hadoop, SQL, Apache Spark.

Work Experience

AIG, Data Science Intern, New York, Summer 2015

- Applied several natural language processing techniques to gather insight from massive collections of insurance claim data.
- Investigated different topic modeling techniques to analyze trending insurance risk topics.
- Worked with different big data platforms such as Hadoop and Apache Spark.

Select Publications

Journal Papers

- H Soleimani, D J Miller, "ATD: Anomalous Topic Discovery in High Dimensional Discrete Data," *IEEE Transactions on Knowledge and Data Engineering*, DOI: 10.1109/TKDE.2016.2561288, preprint: arXiv:1512.06452. Code: https://github.com/hsoleimani/ATD
- H Soleimani, D J Miller, "Parsimonious topic models with salient word discovery," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, pp. 824-837, 2015. preprint: arXiv:1401.6169. Code: https://github.com/hsoleimani/PTM
- D J Miller, H Soleimani, "On an Objective Basis for the Maximum Entropy Principle," *Entropy*, vol. 17, pp. 401-406, 2015.
- H Soleimani-B, C Lucas, B N Araabi, L Schwabe, "Adaptive prediction of epileptic seizures from intracranial recordings," *Biomedical Signal Processing and Control*, vol. 7, pp. 456-464, 2012.
- H Soleimani-B, C Lucas, B N Araabi, "Fast evolving neuro-fuzzy model and its application in online classification and time series prediction," *Pattern Analysis & Applications*, vol. 15, pp. 279-288, 2012.

Conference Papers

- H Soleimani, D J Miller, "ATD: Anomalous Topic Discovery in Text Documents," In *ICML 2016 Anomaly Detection Workshop*, pages 1-5, 2016.
- H Soleimani, D J Miller, "Exploiting the Value of Class Labels in Topic Models for Semi-Supervised Document Classification," *IEEE International Joint Conference on Neural Networks*, 2016. Code: https://github.com/hsoleimani/MCCTM
- H Soleimani, D J Miller, "Sparse Topic Models by Parameter Sharing," In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pp. 1-6.

Select Research Experience

- **Multi-label Document Classification and Credit Attribution**
  Introduced a method for classifying text documents and jointly doing label attribution to individual sentences. Each sentence may explain only a subset of the documents labels, with the label set of the document the union of the labels of all its sentences. Very few previous methods allow joint document labeling and detailed sentence-level attribution.
- **Anomalous Topic Discovery**
  Proposed an algorithm for detecting patterns exhibited by anomalous clusters in high dimensional discrete data such as text documents. Rather than detecting individual anomalies, our proposed method detects groups (clusters) of anomalies (i.e. sets of points which collectively exhibit abnormal patterns) and the highly sparse subsets of the feature space on which the anomalous patterns manifest. Our method demonstrates very promising experimental results compared with alternative techniques for this highly challenging problem.
- **Parsimonious Topic Modeling**
  Proposed a Bayesian Information Criterion (BIC) objective function to exploit parsimony and parameter sharing in topic models for unsupervised processing of massive text corpora. Our proposed model jointly identifies a set of salient words under each topic, determines the small subset of relevant topics for each document, and estimates the total number of topics in the corpus. Our proposed model significantly outperforms other topic models with respect to several clustering performance measures.
- **Maximum Entropy Conditional Probability Modeling for detecting SNP interactions**
  Experimental comparison between maximum entropy models with joint and conditional probability constraints for detecting Single Nucleotide Polymorphism (SNP) interactions.

Teaching Experience

- Lab Instructor, Digital Signal Processing, Pennsylvania State University, 7 semesters.
  - Latest Student Rating of Teaching Effectiveness (SRTE): 6.63/7.0, Average: 5.78.
  - Contributed in designing the lab experiments.
- Invited Speaker, Tutorials on Machine Learning, AIG, summer 2015.
- Teaching Assistant, Linear Control Systems, Ferdowsi University of Mashhad, 1 semester.

Service Activities

Reviewer for IEEE Transactions on Neural Networks and Learning Systems, Journal of Neuroscience Methods, International Machine Learning Conference (ICML), Evolving Systems, ICASSP, and ICDCS.

Select Graduate Courses

Monte Carlo Methods, Dimension Reduction Methods, Statistical Computing, Asymptotic Analysis, Probability Theory, Machine Learning, Stochastic Processes, Neural Networks, Information theory
Independent Coursework: Probabilistic Graphical Models (Coursera - Stanford University), Natural Language Processing (Coursera - Columbia University), Data Analysis (Coursera - Johns Hopkins University)

References

Available upon request.