Sathappan Muthiah

3919, 4th St N, Apt 4,
Arlington, VA-22203
Work Permit: F1 until 2020
E-mail:sathap1@vt.edu
Phone:801-201-0619
Github: https://github.com/sathappanspm
Website: http://people.cs.vt.edu/sathap1/

Summary

I am a Phd student working in applied Natural Language Processing (NLP). My research involves understanding and extracting necessary information from text (obtained from Open source Indicators (OSI) like Twitter, News, Blogs, etc.) with additional understanding of uncertainty and other constraints like minimal coverage, availability of human supervision time etc., for creating cost-effective tabular event data which has applications in event forecasting and healthcare research. My work has led to publications in top-tier data mining conferences like KDD, IAAI, IJCAI, CIKM, etc.

Experienced with working on big data, text and time series mining. Broadly interested in text mining / Applied NLP, machine learning, and data science.

Projected to graduate in Spring 2020

Education

- **PhD** in Computer Science — June 2014 — present
  Virginia Tech — Arlington, VA
  GPA: 3.80
  Advisor: Dr. Naren Ramakrishnan
- **MS.** in Computer Science — Jan 2012 — June 2014
  Virginia Tech — Arlington, VA
  GPA: 3.82
  Advisor: Dr. Naren Ramakrishnan
- **B.Tech. in Information Tech.** — 2007 — 2011
  National Institute of Tech. — Bhopal, India
  GPA: 8.65

Professional Experience

- **Virginia Tech** — Spring 2012 — Present
  GRA, Discovery Analytics Center — *Arlington, VA*
  - **Human-aware Approaches for Event Encoding:** The project is aimed building a cost effective system for encoding political events like Protests, Military Actions etc from news media. The major goal of the system is to create a system with better understanding of uncertainty i.e., it knows when it doesnt know. Characterizing uncertainty helps determine which sections/documents need human supervision. Overall the goal of the project is to maximize coverage via automated approaches for a given accuracy level and at the same time identify documents that can be sent for human supervision (with max available time constraints) to improve overall performance of the hybrid AI + Human system.
  - **EMBERS (IARPA Open Source Indicators program):** Worked as a part of multi-disciplinary multi-university team on building a real-time forecasting system of spatio-temporal events like Civil Unrest. EMBERS was sponsored by a three-year contract for approx. $13.36M from the IARPA OSI Program. Led development of model to identify planned events from social media, news and other sources and approaches to create automated narrative (sequence of) stories from real-time news. Also researched on approaches to perform automated geo-coding (geo-disambiguation) of news and other textual content. Besides contributed in system architecture and ingestion/enrichment of data sources.
  - **SAFE (IARPA Mercury):** Led development of fusion system. The fusion system is responsible for accepting alerts from all underlying models and deciding which alerts (each alert is a tuple of <forecast date, location, actors, event-type>) to merge and which ones to suppress. Built models to perform real-time clustering of alerts and to predict expected quality of an incoming alert which can then be used to merge/suppress alerts.
  - **Time Series Anomaly Detection and Forecasting for Cyber-Physical Systems:** Studied utility of Sequence-to-Sequence (Seq2Seq) models for the task of multi-variate time series state forecasting. Characterized need for changes to traditional seq2seq architecture (developed predominantly for NLP) for applying in time-series and thereby collaborated in effort to build dynamic attention networks (DyAT) for state forecasting.The work is currently under review in IJCAI 2019

– **IGACAT (IARPA Functional Genomic and Computational Assessment of Threats Program):** Researching on methods to characterize threats from specific genes/proteins using heterogenous interaction graph created from pubmed/medline abstracts and information of GO functions, sequence groups and cluster information obtained from knowledge bases like UNIPROT, TREMBL etc. Built deep learning models (CNN, LSTM) for performing few shot learning to identify functions of a nucleotide sequence by embedding both the sequence and label (the functions) in the same space.

- **Indian Institute of Scientific Education and Research** Nov 2008 — Jan 2009
  Intern, advised by Dr. Prashanta Panigrahi *Kolkata, India*

  – Studied the feasibility of a quantum computing system using common and attainable finite dimensional multipartite quantum states.

- **National Institute of Oceanography** May-June 2010
  Intern, mentored by Dr. Biswajit Chakraborty *Goa, India*

  – Estimated geophysical parameters using SONAR Backscatter data obtained off an experiment on the Western Continental shelf of India. Developed processing and cleansing techniques to process backscatter data and used simplex method for estimation.

Publications

Accepted

N. Muralidhar, S. Muthiah, K. Nakayama, N. Ramakrishnan, and R. Sharma. Multivariate long-term state forecasting in cyber-physical systems: A sequence to sequence approach. *Big Data*, 2019

M. R. Islam, S. Muthiah, and N. Ramakrishnan. NActSeer: Predicting User Actions in Social Network using Graph Augmented Recurrent Neural Network. In *Accepted at CIKM 2019*, 2019

M. R. Islam, S. Muthiah, and N. Ramakrishnan. RumorSleuth: Joint Detection of Rumor Veracityand User Stance. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2019

N. Muralidhar, S. Muthiah, and N. Ramakrishnan. Dyat nets: Dynamic attention networks for state forecasting in cyber-physical systems. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 3180–3186. International Joint Conferences on Artificial Intelligence Organization, 7 2019

D. K. Gupta, S. Muthiah, D. Mares, and N. Ramakrishnan. Forecasting Civil Strife: An Emerging Methodology. In *HUSO The Third International Conference on Human and Social Analytics*, 2017

P. Chakraborty, S. Muthiah, R. Tandon, and N. Ramakrishnan. Hierarchical Quickest Change Detection via Surrogates. *arXiv preprint arXiv:1603.09739*, 2016

S. Muthiah, B. Huang, J. Arredondo, D. Mares, L. Getoor, G. Katz, and N. Ramakrishnan. Capturing planned protests from open source indicators. *AI Magazine*, 37(2), 2016

S. Muthiah, P. Butler, R. P. Khandpur, P. Saraf, N. Self, A. Rozovskaya, L. Zhao, J. Cadena, C.-T. Lu, A. Vullikanti, A. Marathe, K. Summers, G. Katz, A. Doyle, J. Arredondo, D. K. Gupta, D. Mares, and N. Ramakrishnan. Embers at 4 years: Experiences operating an open source indicators forecasting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 205–214, New York, NY, USA, 2016. ACM

Y. Ning, S. Muthiah, H. Rangwala, and N. Ramakrishnan. Modeling precursors for event forecasting via nested multi-instance learning. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1095–1104, New York, NY, USA, 2016. ACM

S. Muthiah, B. Huang, J. Arredondo, et al. Planned Protest Modeling in News and Social Media. In *AAAI Conference on Artificial Intelligence, January 25-30, 2015*, pages 3920–3927, 2015

J. Schlachter, A. Ruvinsky, L. A. Reynoso, S. Muthiah, and N. Ramakrishnan. Leveraging topic models to develop metrics for evaluating the quality of narrative threads extracted from news stories. *Procedia Manufacturing*, 3:4028–4035, 2015

Y. Ning, S. Muthiah, R. Tandon, and N. Ramakrishnan. Uncovering news-twitter reciprocity via interaction patterns. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, pages 1–8. IEEE, 2015

N. Ramakrishnan, P. Butler, S. Muthiah, N. Self, et al. 'Beating the news' with EMBERS: Forecasting Civil Unrest using Open Source Indicators. In *International Conference on Knowledge Discovery and Data Mining, KDD, August 24 - 27, 2014*, pages 1799–1808, 2014

A. Doyle, G. Katz, K. Summers, C. Ackermann, I. Zavorin, Z. Lim, S. Muthiah, P. Butler, N. Self, L. Zhao, et al. Forecasting significant societal events using the embers streaming predictive analytics system. *Big Data*, 2(4):185–195, 2014

Y. Ning, S. Muthiah, N. Ramakrishnan, H. Rangwala, and D. Mares. When do crowds turn violent? uncovering triggers from media. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018

M. R. Islam, S. Muthiah, B. Adhikari, B. A. Prakash, and N. Ramakrishnan. Deepdiffuse: Predicting the 'who' and 'when' in cascades. In *2018 IEEE International Conference on Data Mining (ICDM)*, Nov 2018

Activities

- **Intellectual Property:** P. Butler, S. Muthiah, and N. Ramakrishnan. Alert Generation from Multiple Streams, 2016. VTIP 17-047
- **Program Committee:** ACM SIGKDD 2018-19, WWW 2019, ASONAM 2018

Awards and Honors

- **Student Travel Award** 2016
  KDD
- **Pratt Fellowship** 2015
  Virginia Tech
- **Deployed Application Paper Award** 2014
  IAAI

Technical Skills

- **Programming:** Python, C/C++, R, Java, Matlab, HTML, PHP
- **Frameworks:** NOSQL: MongoDB, ElasticSearch, Tensorflow,PyTorch MapReduce

Participation in Conferences

- **KDD 2016**, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, August 2016.
- **KDD 2015**, 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, August 2015.
- **IAAI 2014**, 27th Conference on Innovative Applications of Artificial Intelligence, Austin, USA, Jan 2014.