I am interested in Applied Machine Learning and Data Visualization with special focus on Natural Language Processing, Text Mining, Information Extraction, Social Media Analysis and Deep Learning. I have received international awards for both Machine Learning and Data Visualization. I am also a Full Stack Developer and have an industry experience of four years. I am graduating in May 2018.

Please visit [http://people.cs.vt.edu/parang/category/research](http://people.cs.vt.edu/parang/category/research) for a detailed description of each of the research projects, along with explanatory videos and slides.

- **EMBERS AutoGSR: Semi-Automated Approach for Comprehensive Event Extraction from News Articles**
  
  *Research Areas: Information Extraction, Deep Learning, Natural Language Processing, Visualization*
  
  EBERS AutoGSR is a novel, web-based framework that generates a comprehensive database of validated civil unrest events using minimal human effort. AutoGSR employs an ecosystem of machine learning models along with a carefully designed user interface that allows comprehensive extraction of events. AutoGSR is a deployed system since August 2016 that is continually processing data 24x7 in a semi-automated fashion. The system extracts civil unrest events of type “who protested when, where and why?” from news articles published in over 9 languages, and collected from 18 countries.

- **EMBERS: Civil Unrest Forecasting using Open Source Data**
  
  *Research Areas: Social Media Analysis, Forecasting, Big Data, Intelligence Analysis*
  
  EBERS is a $20M IARPA funded project that is fully automated, cloud hosted, big data system that generates forecasts for significant societal events using open source data including Twitter, Facebook, RSS sources, News, blogs, Google search trends, meteorological data, economic and financial indicators, satellite imagery etc. EBERS processes massive data streams in real time and generates alerts for significant societal events.

- **DMAP: Data Aggregation and Presentation Framework**
  
  *Research Areas: Natural Language Processing, News Visualization*
  
  DMAP (Data Mining and Automation for Platforms) is an online framework that presents a wide variety of data, news and information about companies. It is developed with an aim to act as a one-stop platform for displaying everything official related to a company and its competitors. It aggregates data from a wide range of data sources and displays them in an intuitive fashion that allows a user to perform exploratory analysis.

- **Forex-Foreteller: Currency Exchange Rates Determiner**
  
  *Research Areas: Financial Modeling, Prediction, Trend Analysis, Visualization*
  
  Forex-foreteller mines news articles and makes forecasts about the movement of foreign currency markets. The system uses a combination of language models, topic clustering, and sentiment analysis to identify relevant news articles. These articles along with the historical stock index and currency exchange values are used in a linear regression model to make forecasts about the movement of currency exchange rates.

- **Epidemiological Modeling of News and Rumors on Twitter**
  
  *Research Areas: Network Analysis, Social Media Analysis, Epidemiological Modeling*
  
  Characterizing information diffusion on social platforms like Twitter enables us to understand the properties of underlying media and model communication patterns. As Twitter gains in popularity, it has also become a venue to broadcast rumors and misinformation. We use epidemiological models to characterize information cascades in twitter resulting from both news and rumors.
Please visit [http://people.cs.vt.edu/parang/category/visualizations](http://people.cs.vt.edu/parang/category/visualizations) for a detailed description of each of the Visual Frameworks, along with explanatory videos and slides.

- **Semi-Automated Extraction of Events from News Articles**
  This framework was developed in conjunction with the EMBERS AutoGSR and provides a visual interface that allows extraction of events from news articles while minimizing the manual effort required for doing so. The system shows both full and partial event recommendations in order to assist users with event extraction.

- **Visualizing Civil Unrest Events of Type, ‘Who Protested, When, Where and Why’**
  This framework employs coordinated charts that allow analysts to club multiple query parameters in order to visualize spatio-temporal events. Using this visualizer, analysts can uncover patterns and associations between named entities and can perform both macro and micro level exploratory analysis of the extracted events.

- **Exploratory and Comparative Analysis of Organizations**
  This framework was developed as a part of the DMAP system and acts as a one-stop platform for analyzing wide variety of information associated with companies. The interface provides an intuitive visualization of news articles along with extracted entities that allows for easy discovery of connections between entities.

- **News Analyzer**
  The News Analyzer interface provides novel ways to visualize and analyze news articles collected from different sources. An analyst can identify trending entities like name of people, organizations and locations in the news articles and can also see how the corresponding trends have evolved over time.

- **Email and Network Analyzer**
  The Email Analyzer interface helps an analyst visualize the email network and identify local group of people who frequently exchange emails amongst themselves. The interface uses a radial network visualizer along with a dynamic scatter plot to identify groups of users, who most frequently exchange emails amongst themselves.

- **Geo-Spatial Data Analyzer**
  The Geo-Spatial interface provides intuitive ways of visualizing and analyzing GPS coordinates of users. The interface allows for playback of GPS locations of selected users over the map, thereby making it easy to identify locations, co-located users, and most favored routes.

- **Streaming Social Media Data Analyzer**
  The interface for Streaming Data Analyzer allows for visualization of streaming data from Twitter. It displays the currently trending keywords along with the evolution of trends over time. Analyst can also visualize geo-located tweets on map.

**Past Employment**

- **HP Patent Research, CPA Global, India**
  *Information Retrieval, Narrative Generation*
  Our division was responsible for performing patentability searches, validity/invalidity searches, infringement analysis, landscape searches, accelerated examination searches and identification of evidence of usage.
  - I was responsible for developing a document retrieval system that could fetch relevant studies based on similarity of content from a database of prior-art studies performed in the past by our division. This also included developing web-based interface for populating the database by other research analysts.
  - I also developed an intelligent report generating software that could automatically fetch meta-information from online web sources for patents and academic publications and populate well-formatted reports to be delivered to the client.

- **Corporate Executive Board (CEB), GlobalLogic India Ltd.**
  *Recommendation Systems, Full-Stack Web Development*
  Our division was responsible for designing and developing a 4-tier web architecture that was flexible enough to power all the 60+ websites of CEB and still provide enough functionality to modify individual websites.
  - I was one of the core Full-Stack developers who worked on both frontend and backend of this web framework.
  - I was also responsible for writing a recommendation module that would suggest case studies and industry reports to the users of CEB based on their past purchases, preferences, document similarity and user similarity.
Please visit [http://people.cs.vt.edu/parang/category/publications](http://people.cs.vt.edu/parang/category/publications) for a detailed description of each of the publications, along with explanatory slides, posters, videos and full paper.

**Under Review**


**SIGKDD**

- Sathappan Muthiah, Patrick Butler, Rupinder Paul Khandpur, Parang Saraf, Nathan Self et al. “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, 2016. (Acceptance Rate: 20%)

**ECML PKDD**


**TECHNOMETRICS**


**IEEE VIS**

Achievements

International Awards
• 2016 Jack Youden Award for the best expository paper from the American Society for Quality for my paper on Bayesian Model Fusion for Forecasting Civil Unrest that appeared in Technometrics 2015
• Winner – 2014 IEEE VIS VAST Challenge
• Best Paper Award for Epidemiological Modeling of News and Rumors on Twitter at 7th SNA-KDD 2013

Fellowship(s)
• Pratt Fellowship (2013) in recognition of outstanding record as a Graduate student in Computer Science

Invited Talks
• Pecha Kucha Talk on EMBERS at 2015 Big Boulder Conference, Boulder, CO
• Talk on News Analyzer Framework at 2015 Industries Studies Association, Kansas City, MO

Travel Awards
• AAAI 2018 Student Travel Award
• SIGKDD 2016 Student Travel Award
• IEEE VIS 2014 Student Travel Award
• SIGKDD 2013 Student Travel Award

Research Achievements
• Received $1M funding from IARPA to extend my Ph.D. dissertation on comprehensive event extraction to new event types and languages
• Core member of the EMBERS team that won the IARPA OSI challenge for forecasting significant societal events

Professional Service

Department Editor
• ACM XRDS – Crossroads, Magazine

Program Committee Member
• WSDM 2018
• IAAI 2018
• ACM HyperText 2018

Reviewer
• CHI 2018
• OZCHI 2017
• IEEE VIS VAST Challenge 2017
• ACM Transactions on Interactive Intelligent Systems (TIIS)
• Elsevier Journal on Information and Management (Impact Factor: 3.3)

Session Chair
• IAAI 2018 – Session on Supporting Government Services

Technical Skills
• Python, C, Shell Script, R
• HTML, CSS, Bootstrap, JavaScript, jQuery
• D3, Crossfilter
• Django
• SQL, PostGRE, SQLite

Last Updated: March 12th, 2018