Mayank Gupta

iamanugly@duck.com bayesrule.me

EDUCATION

MS in Computer Science — New Brunswick, NJ, USA | May 2016
RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

B.Tech in Information Technology — Vellore, TN, India | May 2012
VELLORE INSTITUTE OF TECHNOLOGY

MASTER'S THESIS

COMPARISON OF THE TRIANGLE ALGORITHM AND SMO FOR SOLVING THE HARD MARGIN PROBLEM

Rutgers University

Implemented the algorithm for the hard-margin case of *support-vector machines*, improved computation efficiency, contributed to errata, contributed theoretical ideas and developed a web UI to visualise results in 2D and various matlab visualisation for higher dimensional data.
Advisor: Bahman Kalantari, Dept. of Computer Science, Rutgers University

WORK EXPERIENCE

BLOOMBERG LP | Software Engineer — Skillman, NJ/NYC, NY | July 2016 – Present

- DTCommunity Profiles (July 2016- August 2016)
  - *MVP* - Designed and implemented a real-time mechanism to update user profile views on MVP. The existing solution updated it every hour. DMP + Kafka + CPP + Comdb2 + Rapid
  - Worked on various fixes and small features. eg: Index additional content to their SOLR database, worked with legacy databases, migration of legacy services. CPP + SOLR + Shell scripts
  - Used a *topic-modelling algorithm* to analyze user status updates and extract user interests. Labelled-Latent dirichlet allocation, Python
  - Used the association rule mining algorithm to post-process user interests to make meaningful recommendations.
- New hire training (August 2016- November 2016) - Overview of bloomberg technology stack.
- DT Company Financials (Dec 2016- April 2019)
  - *FEDSVC* - Designed and implemented a language-independent model based on graph theory to process the excel models for bank analysts publishing company reports. This model is used everyday to feed data to the MODL screen with a daily report on the performance. Cython + Python + Algorithms + BCOS
  - The algorithm is generic and used to process tables for edgar filings and any table type in general.
  - Researched theoretical bounds for the algorithm to prove its optimal and cannot be improved. The existing solution used a language dependent clustering algorithm with precision/recall numbers in the ballpark of 80%. With the graph theory approach we achieved 100% precision and recall while keeping the processing time to below 2secs for most cases.
  - *Anchors* - Developed a O(nlogn) algorithm to assert uniqueness of data in the excel models to improve entity mapping. This algorithm saved about 2 hours of manual work per model and was added to FEDSVC to achieve end-to-end processing with no human intervention.
  - *Nextgen and Deep estimates* - Designed and implemented workflows to integrate FEDSVC for company filings and analyst estimate. The workflows used various idempotent services to avoid various issues with state management. DTP + BBDS + kafka
  - *DTQ* - Contributed to a lockless queue with comdb2 to process the traffic for the above workflows. The lockless queue used a test and set primitive to avoid holding a lock. The queue supports throttling and priority of requests. Comdb2 + Python
  - *CODFOA* - Contributed to a fast CPP + Cython library that provides a common format to convert various natural language documents. Fixed bugs in the streaming version of the library. Cython + CPP
  - Worked on various other fixes and improvement with the team to improve interfacing with the systems and general stability.

- *Grumpy* - Developed a self-supervised learning algorithm as a low cost word segmentation fix for various errors related to the existing optical character recognition solution. Python
- Developed numerous prototype web UIs to debug and analyze the results of algorithms.
- Contributed to various dashboard to *collect metrics* and display aggregated results to monitor system health and analyze the impact of software changes.
- Worked on numerous *performance and memory improvements* to keep the response times within an acceptable range during the course of development/upgrades.
- Used metric/distance learning to learn local word embedding for analyst documents and used MDS algorithm to visualize the results in 2-D. The embedding were effective in learning meaningful representations including between abbreviations and their full-form.

- DT Automation (April 2019 - April 2020)
  - *Table extraction* - Designed and developed a O(nlogn) language-independent algorithm for finding columns in a table based on *computational geometry*. The existing solution used a random forest, was language dependent and ran on specialized hardware. Blockdoc + CPP + Cython + Python
  - *Table extraction* - Designed and developed a O(N) language-independent algorithm to detect start and end of tables with column layout. Coupled with the above algorithm for column finding my solution extracts tabular structure for most documents without any specialized hardware that is required for running neural networks. Blockdoc + CPP + Cython + Python
  - *Table row hierarchy detection* - Designed and developed a O(N)-amortized language independent algorithm for detecting relationships between various rows of a table. My solution could stay language-independent because instead of looking at the content, it looked at the style of each text like stroke, color or font. Reduced the Anchor algorithm from my previous team to a special case of this new algorithm with reduced time complexity.
  - *APIs and fixes* - Did some fixes to the code for existing table extraction library, identified limitations in the current deep-learning based solution, wrote some web-assembly and cython code for compatibility across platforms, fixed memory issues and bugs in internal libraries, implemented a workflow for backfilling of data and developed splunk dashboard to view service metrics and fixed issue with the random forest classifier.
- DT Quality (July 2020 - June 2023)
  - *APIs and fixes* - Worked on feature improvements and fixes to the existing distributed system solution to generate quality metrics for data processing pipelines. Kafka + Hive + Comdb2 + Python
  - *Random Sampling* - Worked on various feature improvements, developing and deploying their react website with form validations and web-analytics. Python + react + BPaaS + Matomo
  - *Random Sampling* - Contributed to theoretical ideas to reduce manual effort required for processing.
  - *Provenance* - Worked on some UI improvements and stability fixes for a workflow provenance tool. SQL + Python + Javascript
  - Worked on *performance improvements* for effective resource utilization and developed *splunk dashboards* to monitor improvements.
  - Developed prototype workflow to demonstrate system capabilities and use as a template for similar tasks.
- AIM Compliance(July 2023 - Present)
  - Advocated for the use of IO/CPU efficient algorithms for various data retrieval and machine learning tasks.
  - Documented the design for legacy code base and the suggested refactoring based on contemporary design principles and technologies.

RUTGERS UNIVERSITY | Graduate Assistant (GA) Piscataway, NJ | Nov 2014 - May 2015

- Worked at Center for Advanced Infrastructure and Transportation.
- Developed a *web based GIS system* and various other visualization methods to display geographical data.
- Developed a *3D web based visualiser* to view the structure and health of bridges using WebGL and three.js.
- Developed multiple pages to visualise data using charts and tables.
- Worked on *performance improvements* to keep the UI responsive with 150,000 locations visualized.
- The tool beat various state of the art systems by leveraging the recent advances in *WebGL and Maps*.

BLOOMBERG LP | Summer Intern Skillman, NJ | June 2015 – August 2015

- Developed web UI and services to process *social graphs for the bloomberg user and financial network*. Neo4j + Python + Java + Javascript

- Wrote a custom plugin for Neo4j *path finding algorithms*. Java
- Developed a *combinatorial algorithm* to process simple user query in natural language and added a TRIE based *autocomplete*. Python + Javascript
- Analyzed social connections by running *clique detection* and *page rank*.
- Overlayed bloomberg news articles over profiles using the web UI to analyze social relationships.

SOURCEBITS | Software Eng/Game developer — Bangalore, KA | July 2012 - May 2014

- Worked on the game engine on an Ipad based *strategy game*. The game engine was built in-house so a lot of work went into fundamental algorithms. C++ OpenGL
- Made critical fixes and enhancement to the path-finding algorithm.
- Implemented various models of combat behavior in the AI engine (troop movements, weapon types, border wall behavior, aerial units etc).
- Implemented various visual effects of combat behavior in the graphics engine (camera movement, health banners, heart palpitations, weapon effects etc)
- Fixed some floating point issues related to space discretization in the physics engine.
- Stress tested the engine and did various performance enhancement until the desired level.
- Developed the backed services for an *Ipad media application*. Web scraping + NodeJS + Amazon AWS + Content indexing + load balancing + RSS feeds.
- Achieved efficient resource utilisation for the web scraper using an asynchronous framework.
- Exposed REST APIs using NodeJS, MongoDB and Apache SOLR.
- Stress tested the system and worked on numerous *memory and performance fixes*.

PIVOTAL SECURITIES | Summer intern — Vellore, TN | May 2011 - July 2011

- *Security Integration and Event Management* - Worked on the prototype product for a startup.
- Extended the *rule inference engine* CLIPS (from NASA) to provide database querying capabilities. CLIPS + MongoDB
- Wrote prototype rules in CLIPS to perform various inferences related to the domain.
- Implemented a client-server script using *OpenSSL* C API to collect various logs from machines, normalise and store in the database for rule inference. C, OpenSSL

MISCELLANEOUS

DUCKDUCKGO | Zero-click info plugin | Aug 2012- Aug 2012

Wrote a plugin that displays an arbitrary-funny joke as the first result in the search engine when searched for keywords like *"Chuck norris jokes/facts"*. The plugin is now deprecated and replaced by wikipedia info. Javascript/Perl

ONLINE COURSES | MOOC | Aug 2011 - Present

List of notable courses completed with certifications

- Deep learning specialization - Coursera
- Artificial Intelligence - BerkeleyX (edX)
- Generative Adverserial Networks (GANs) specialization - Coursera
- Financial Markets - Robert Shiller - Coursera

Other unfinished course materials and personal projects to be found at my code repository.

CUBESAT | University small satellite program | Aug 2010 - Aug 2011

Participated in the program for developing a 10x10x10cm *cubical satellite* with the aim of taking photographs in orbit. My job was to write up the technical specifications for the onboard compute module, familiarise with the other components and review the bill of materials. The goal did not materialize but our team presented a report to a team from the *Indian Space Research Organization*.