sdmay23-35

Learning the Popularity Prediction in Information Cascades

Faculty Advisor & Client: Goce Trajcevski

Paul Brinkmann | Bailey Gorlewski | Evan Gossling | Ian Johnson | Will Postler

Intro & Overview

Many researchers work on addressing the problem of information cascades using the APS data set.

The objective of this project is to develop an end-to-end system that will provide cascade prediction and visualization functionalities for the users interested in visualizing locations related to popularity cascades in the context of scientific paper citations.

Intended Users

- Individual Scientist
- Funding Institution
- University Administrator

HINTS ML Model

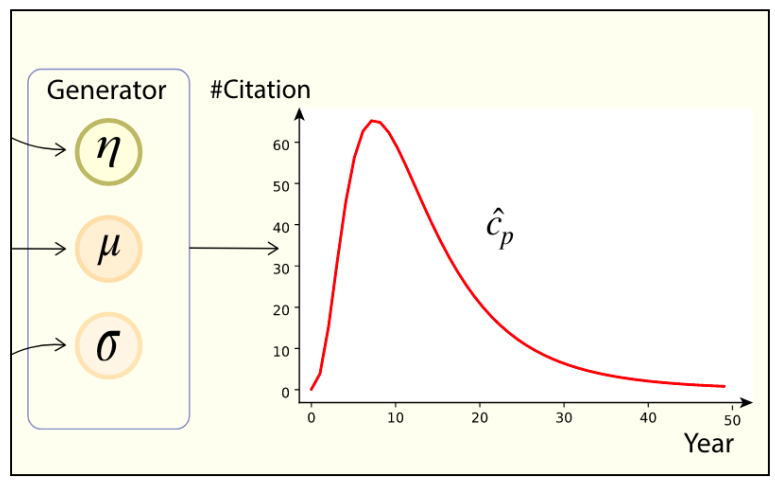
The "Generator" function utilizes η, μ, σ which are obtained via the Machine Learning Model. These values are then plugged into the following equations to obtain our graph which is displayed to our users

Integrated across η_p, the predicted cumulative citation counts $\hat{C}_p^l$ of paper p at l-th year after publication can be generated by

$$\hat{C}_p^l = \alpha\left[\exp\left(\eta_p * \Phi\left(\frac{\ln l - \mu_p}{\sigma_p}\right)\right) - 1\right] \quad (6)$$

where $\Phi(x)$ is

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-y^2/2}\, dy. \quad (7)$$



Standards

- IEEE 830 - Software Requirements Specifications
- IEEE 1016 - Software Design Descriptions
- IEEE 12207 - Software Life Cycle Processes
- IEEE 1028 - Software Reviews and Audits

Implementation

- Python
- Flask
- HTML/Javascript/CSS
- JS Libraries (jQuery, Ajax)
- Mapbox
- MySQL
- Docker
- CI/CD

Results

We were able to successfully build an end-to-end system which takes a paper from the APS dataset and generates graph and map visualizations based on the ML model's results very fast. Product was verified by our client.

Impact

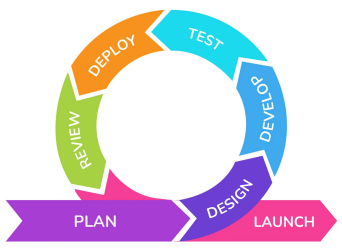
The broader impact is that we have successfully implemented the HINTS ML Model into a web application allowing for easy access for our intended users and use cases.

Future Work

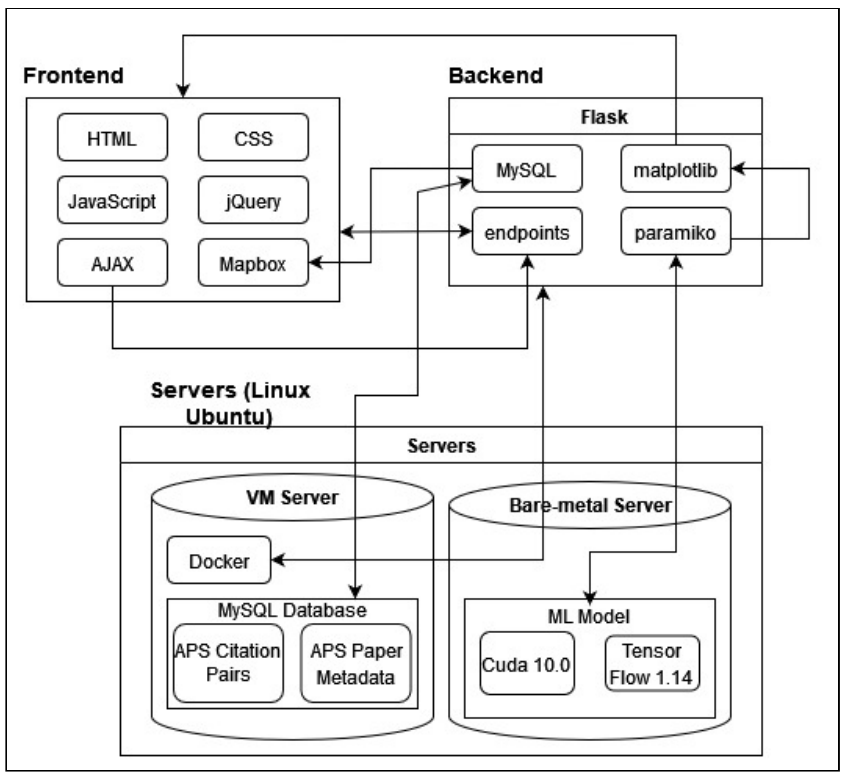
Future work involves a better implementation/integration of the ML model to better customize the visualization capabilities.

Methodology

- Lotus Blossom
- Decision Matrixes
- Agile Development
- Client Validation



System Architecture



Testing

- PyTest
- Integration Testing
- Interface Testing
- Regression Testing
- Acceptance Testing

Challenges

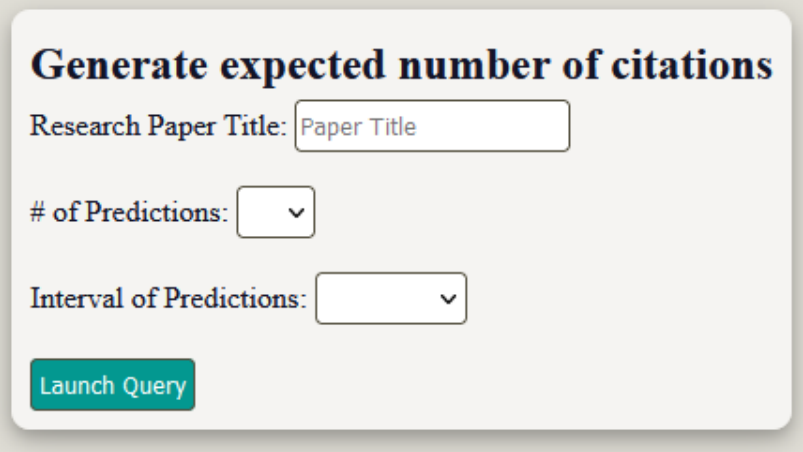
- Usage of React
- Model required Cuda 10.0
- Dataset format

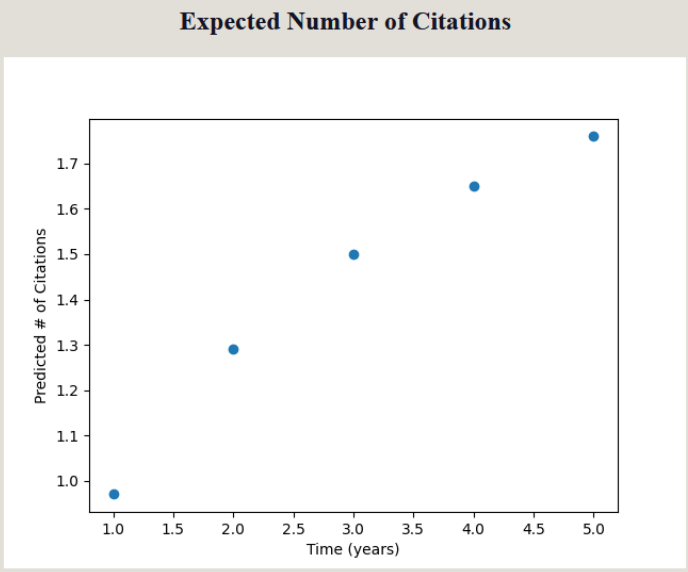
Solutions

- Vanilla JS
- Acquired physical server
- Custom import script

Conclusion

In conclusion we developed an end-to-end visualization web application for prediction information cascades for research paper citations. We utilized a Flask backend, with a vanilla JS frontend that used the Mapbox API in order to create geospatial visualizations of paper citations, and we generated visualizations of cascade predictions





Please see our final report document for more information!