

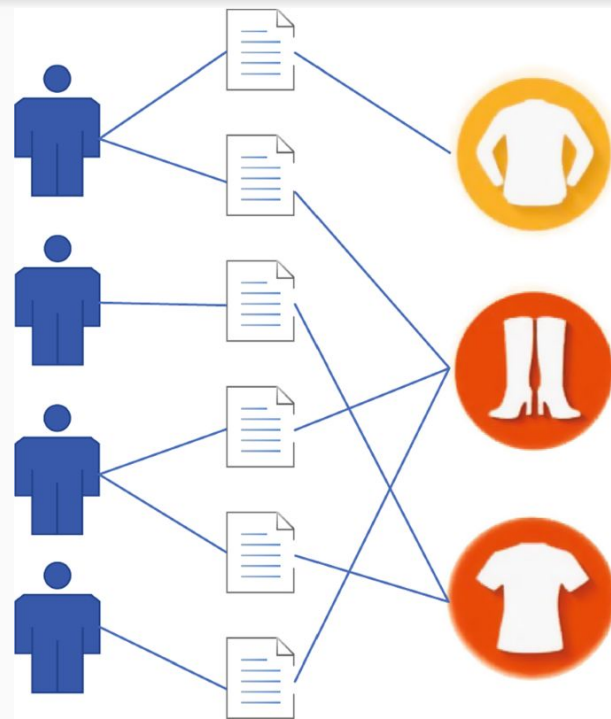
Human-In-The-Loop Capstone Final Presentation

Noah Backman, Jinan Hong, Andy Kelly, Chris Munoz



Background & Overview

- Professor Xie
- Spam Detection
 - Yelp, Amazon
- Machine Learning
 - SpEagle
 - MRF
- Many traits of a review
- Website



Start of Capstone Experience

- Unfamiliar with this level of ML
- Inherited code
- Foundations of website
- A focus on functionality and purpose



First Semester Summary

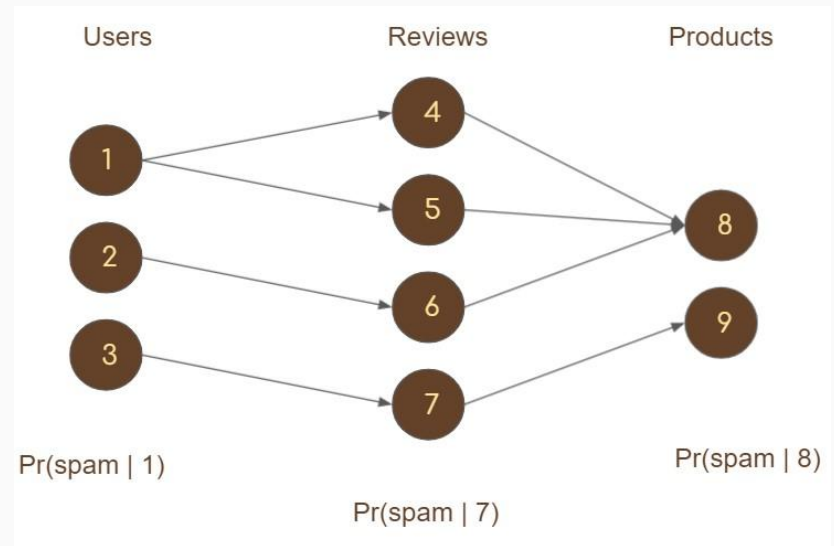
- Lot's of ML and webdev research
- Lot's of debugging inherited code
 - Routing, passwords, missing pages, documentation
- Hosted site on Lehigh server
- Built and connected live mySQL database

Second Semester Plans

- Bigger focus on SpEagle and ML
- Making the site more interactive and user friendly/educational
- Connect the model and its results to the site
- General aesthetic improvements of the site
- Document our work/codebase to better prepare the next group

SpEagle

- Markov Random Field
- Uses many features of users and reviews
- Guilty-by-association
- Percentile Rankings



Interpreting SpEagle

Model demo

Website

- Great improvements to Training/Testing pages
- Added Sandbox page and 2 connected pages for model results
- Overall improvements to site look, flow, and functionality

Live Demo

[Link to the site](#)

Chris

- ML Research
- Front End Improvements
- Research Page
- Sandbox Functionality and HTML

Sort Highest

Sort Lowest

Sort User ID

Enter New Product ID

SpEagle Output for Reviews on Product 12

[Explore](#) USER ID: 562 - PERCENTILE RANK: 46.1538

[Explore](#) USER ID: 563 - PERCENTILE RANK: 53.8462

[Explore](#) USER ID: 564 - PERCENTILE RANK: 30.7692

[Explore](#) USER ID: 565 - PERCENTILE RANK: 7.6923

[Explore](#) USER ID: 566 - PERCENTILE RANK: 15.3846

[Explore](#) USER ID: 567 - PERCENTILE RANK: 38.4615

[Explore](#) USER ID: 568 - PERCENTILE RANK: 76.9231

[Explore](#) USER ID: 4354 - PERCENTILE RANK: 23.0769

[Explore](#) USER ID: 4830 - PERCENTILE RANK: 61.5385

[Explore](#) USER ID: 4831 - PERCENTILE RANK: 84.6154

[Explore](#) USER ID: 4832 - PERCENTILE RANK: 69.2308

Jinan

- ML Research
- Sandbox Work
 - Product Review
- Frontend Formatting

Please input a product ID

Run SpEagle

View All Reviews

SpEagle Output for Product 1

Feature	Value	Baseline	Importance
MNR	0.25	0.596	Multiple reviews in a short time-span in suspicious
PR	0.286	0.503	Spammers sometimes only give high ratings to boost a product
NR	0.408	0.506	Spammers sometimes only give low ratings to tank a product
avgRD	0.934	0.498	Ratings far from the average can be suspicious
ERD	1.972	0.501	Low entropy is suspicious due to lack of expected variation
ETG	2.094	0.493	Low entropy is suspicious as spammers leave reviews regularly

Andy

- ML research
- General site improvements
- Training & Testing pages
- Model connection to site
- Improve readability of sandbox

What do you think SpEagle will predict?

You should use what you learned on the Examples and Training pages to determine if this review is spam.

#	Review Features	Values	Weights
1	Rank of the review	2	.5
2	Extremity of the rating	1	.1
3	Written early?	1	5
4	User's only rating?	0	70
5	Rating deviation from average	1	100

#	Author Features	Values	Weights
1	Most reviews in a single day	3	2
2	Ratio of positive reviews (4-5 star)	13	6
3	Ratio of negative reviews (1-2 star)	0	2.5
4	Standard rating deviation of reviews written	14	2.5
5	Long time between first and last reviews?	2.53	5

Noah

- ML research
- General site improvements
- Model connection to site
- Deployment and Utility Scripts

Reflecting on CSE 280-281

- Learned an immense amount
 - Technical skills: Python, Flask, HTML, CSS, JS, Bootstrap, MySQL and lots more
 - Soft skills: Teamwork, Agile, project/deadline management, communication, etc
- Writing clean, documented good is nearly as important as writing efficient, optimized code
- There are endless possibilities of what you can do with machine learning
- A group's performance will be better when members are friendly and having fun

With more time...

- Continue to add to and improve the site
- Promote the site and research when appropriate
- Continue development of Training & Testing pages
- Browser extension

Thanks for a great capstone experience!

