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
Jinan

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

### SpEagle Output for Product 1

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
<th>Baseline</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNB</td>
<td>0.25</td>
<td>0.596</td>
<td>Multiple reviews in a short time-span in suspicious</td>
</tr>
<tr>
<td>PR</td>
<td>0.286</td>
<td>0.503</td>
<td>Spammers sometimes only give high ratings to boost a product</td>
</tr>
<tr>
<td>NR</td>
<td>0.408</td>
<td>0.506</td>
<td>Spammers sometimes only give low ratings to tank a product</td>
</tr>
<tr>
<td>avgRD</td>
<td>0.934</td>
<td>0.496</td>
<td>Ratings far from the average can be suspicious</td>
</tr>
<tr>
<td>ERD</td>
<td>1.972</td>
<td>0.501</td>
<td>Low entropy is suspicious due to lack of expected variation</td>
</tr>
<tr>
<td>ETG</td>
<td>2.094</td>
<td>0.493</td>
<td>Low entropy is suspicious as spammers leave reviews regularly</td>
</tr>
</tbody>
</table>
Andy

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