Step 4: Project Proposal

Bazeley, Mikiko

Requirements

1. Write your proposal in a Google Doc (1-2 pages) and submit the link via the “Submit” button. Make sure your mentor has the permissions to comment on the document.
2. Work with your mentor to incorporate any feedback into later drafts. You are welcome to submit your proposal as many times as needed.
3. Once your mentor has approved your proposal, convert the doc to a PDF file.
4. Create a GitHub repository for this project (if you haven’t done so already).
5. Add the PDF to your GitHub repository for this project.
6. Share the proposal with your peer community for feedback.

Note: All code and further documentation you write will be added to this repository.

Important Links

- Project Github repo: https://github.com/MMBazel/Wardrobe-Recommender
- Project Documents/Milestones: https://github.com/MMBazel/Wardrobe-Recommender/tree/master/references/project-milestones
- Detailed Working Paper & Notes: [Springboard][MLE] Capstone Project Working Doc- Bazeley, Mikiko
- Slim Notes: [Springboard][MLE][Unit 8.6] Slim List - Bazeley, Mikiko

Proposal

1. What is the problem you want to solve? Why is it an interesting problem?

The importance of fashion to the global economy:

- *" The fashion industry occupies a significant position in the global economy and involves large industrial chains, including garment design, production, and sales. In fact, in the recent years, there has been an expanding demand for clothing all over the world.* ***Since 2008, the garment sales have increased by $3.3 billion every year, and the global garment sales reached $1.25 trillion in 2012*** *[1]. According to a report of Euromonitor International, in 2015, the growth rate of clothing sales was 4.5%, and the industry gross reached $1.6 trillion.* ***The global clothing sales enhanced by 3.8% and the industry gross rose to $1.7 trillion in 2016.*** *The above data show that the garment industry is developing at a rapid rate."* https://doi.org/10.1155/2017/8093057

The fashion industries impact on the environment:

- *"The* ***fashion industry*** *is responsible for* ***10 % of annual global carbon*** *emissions, more than all international flights and maritime shipping combined. At this pace, the* ***fashion industry's greenhouse gas emissions will surge more than 50 % by 2030****."* https://www.worldbank.org/en/news/feature/2019/09/23/costo-moda-medio-ambiente
- *"The fashion industry emits* ***more carbon*** *than* ***international flights and maritime shipping combined.****"* https://www.businessinsider.com/fast-fashion-environmental-impact-pollution-emissions-waste-water-2019-10

More isn't better but maybe it's something else:

- *"Current psychological theory and research affirm the positive affective and motivational consequences of having personal choice. These findings have led to the* ***popular notion that the more choice, the better--that the human ability to manage, and the human desire for, choice is unlimited****. Findings from 3 experimental studies starkly challenge this implicit assumption that having more choices is necessarily more intrinsically motivating than having fewer. These experiments, which were conducted in both field and laboratory settings,* ***show that people are more likely to purchase gourmet jams or chocolates or to undertake optional class essay assignments when offered a limited array*** *of 6 choices rather than a more extensive array of 24 or 30 choices. Moreover,* ***participants actually reported greater subsequent satisfaction with their selections and wrote better essays when their original set of options had been limited.*** *Implications for future research are discussed. " -* *When Choice is Demotivating: Can One Desire Too Much of a Good Thing?*
- *"It's not information overload. It's filter failure"* - Clay Shirky

And the rapid growth and expanse of fashion won't be slowing down anytime soon. Fashion (and fashion brands) found new life in leveraging social media & influencers to push sales by capitalizing on classic psychological tactics surrounding status & belonging:

- "*Our results suggest that **customer participation in a firm's social media efforts** leads to an **increase in the frequency of customer visits**. We find that this participation effect is greater when there are high levels of activity in the social media site and f**or customers who exhibit a strong patronage with the firm, buy premium products, and exhibit lower levels of buying focus and deal sensitivity***." - https://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0460

- *"Findings indicate that **there is a significant relationship between consumers' attitude toward social media advertising and their behavioural responses**. Again, it was noted that corporate reputation moderates this relationship." -* https://www.emerald.com/insight/content/doi/10.1108/JRIM-01-2015-0012/full/html?fullSc=1&mbSc=1

- *"The result indicated that the online marketing communications, specifically, E- WOM, online communities and online advertisement are effective in promoting brand loyalty and product purchase intention through company website and social media platforms." - https://www.sciencedirect.com/science/article/pii/S1877042814039366*

- *"Beyond feelings, people who speak of FOMO also speak of it as **a behavior, most often as a compulsivity (related to what I characterize as conspicuous sociality) and as an illness to be remedied**. And although FOMO is often seen as a recent phenomenon, I argue it is a **continuation of a centuries-old concern** and discourse about media-prompted envy and anxiety (i.e., "keeping up with the Joneses" and neurasthenia)." - https://journals.uic.edu/ojs/index.php/fm/article/view/6064*

- "*We compare, and improve, by way of consumption" -* https://www.fashionrevolution.org/consumption-in-the-digital-age/

- *"Brands that are relegated to following the pack will quickly become irrelevant and will be replaced with zeitgeist-defining brands. How do we know this? Because of the consumer signals we receive. One of those signals, FOMO (a.k.a. fear of missing out) is what's driving Gen Zers — more specifically, the fear of not being relevant because they've missed out on something culturally important. This is the feeling that will completely change the way brands — and even entire markets — operate in the future." -* https://jingdaily.com/why-fomo-is-shaping-the-luxury-market-now-more-than-ever/

Fashion is a big player in the global economy, employing a vast network of individuals and companies that span a complicated supply chain that touches every country in the world, from textile production and dyeing operations in India to factories in China to warehouse operations and shops in Paris, London, New York and San Francisco.

Fashion is also wreaking havoc on the environment due to its heavy resources consumption

and “take, make, dispose” linear economy. Clothing created once is worn for some period of time and then tossed (or a small percentage are recycled), spending years in the landfills leaking out chemicals and spreading microplastics.

And the global appetite won’t be slowing down anytime soon (or ever) as fashion brands have found a new way to market new products and garner brand loyalty through social media advertising and influencer collaborations. By capitalizing on FOMO (which in part is a modern incarnation of the desire for belonging and status), brands push new products knowing that their customers will participate in purchasing future collections by purging their existing closets and sending clothes straight to the trash, regardless of the actual remaining lifetime.

On an individual level, wardrobe optimization can be a confusing and expensive experiment as consumers are exposed to an overwhelming amount of information everyday from a variety of channels.

Thoughtfully designed computer vision recommendation systems can help alleviate the information overload by improving a user’s ability to sort through thrift or second-hand websites and recommend items that fit a person’s existing tastes. By combining the recommendation powers of Amazon and assisting users with finding the perfect thrifted or pre-owned item, we can increase the reuse and longevity of clothing.

2. What data are you going to use to solve this problem? How will you acquire this data?

I will be using three datasets:

1. Large-scale Fashion (DeepFashion) Database - Downloadable through a link to the google drive folder
    a. Description: Large-scale clothes database which contains over 800,000 diverse fashion images ranging from well-posed shop images to unconstrained consumer photos. DeepFashion is annotated with rich information of clothing items. Each image in this dataset is labeled with 50 categories, 1,000descriptive attributes, bounding box and clothing landmarks. DeepFashion also contains over 300,000 cross-pose/cross-domain image pairs.

2. PaperDoll Raw Dataset - Downloadable through link
    a. The Paper Doll dataset is a large collection of tagged fashion pictures with no manual annotation. It contains over 1 million pictures from chictopia.com with associated metadata tags denoting characteristics such as color, clothing item, or occasion.

3. Pinterest Fashion Compatibility - Images need to be called through request
    a. This dataset contains images (scenes) containing fashion products, which are labeled with bounding boxes and links to the corresponding products.
        i. Basic statistics
            1. Scenes: 47,739

2. Products: 38,111
3. Scene-Product Pairs: 93,274

4. Scraped data from my pinterest boards
    a. I have a bunch of pinterest dashboards with images collected over the years, with different items curated in collections of different flavors.
    b. Using a scraper like https://github.com/xjdeng/pinterest-image-scraper, I can scrape my own dashboards to determine what items would be recommended for each collection.

3. In brief, outline your approach to solving this problem. You might not know everything in advance, and this approach may change later. This might include information like:

a. Is this a supervised or unsupervised problem?
b. If supervised, is it a classification or regression problem?
c. What are you trying to predict?
d. What will you use as predictors?
e. Will you try a more "traditional" machine learning approach, a deep learning approach, or both?

The goal is to create a recommendation system based on image. If time permits after the program (or near the end) I may explore adding textual data. (https://openreview.net/forum?id=ryTYxh5ll)

Steps:
1. Collect data
2. Prep: Normalize images
3. Prep: Convert images to embeddings
4. Model: Use Transfer learning from ResNet50 or ResNet34
5. Model: Use PyTorch-Lightning or Keras to retrieve image embeddings
6. Search: Explore using various nearest neighbors algorithms (ANNOY, etc) to obtain most similar images
7. Deploy: Once model code has been created, deploy model using AWS, Docker, Streamlit
8. Evaluation: Evaluation will be manual & will be conducted by manually uploading images from the Pinterest scraper (from different boards) to mimic "user personas". If time permits, I'll do 20 runs and find a way to measure some of the metrics captured in Evaluation Metrics for Recommender Systems.

4. What will be your final deliverable? Will it be an application deployed as a web service with an API or a more robust web/mobile app.

Final deliverables include:

- **Web App** - Images can be uploaded and recommendations in the form of images will be produced
- **Github repo** - Model & code can be downloaded, with instructions on how to get the data
- **Video/screenshare** - I don't intend to host the project forever so instead of someone going to an empty website, I'll have a video walking through the web app
- **Blog Post on Medium** - Walk through of the project

5. What computational resources would you need at a minimum to do this project? *You may not have a very clear sense now, but work with your mentor to come to an estimate. In real industry applications, you'll often be called upon to provide resource estimates at the beginning of a project.*

a. Processing power (CPU)

b. Memory

c. Specialized hardware such as GPUs

I won't necessarily be re-training any deep learning models, instead will be relying on transfer learning, freezing and utilizing the last layers.

I will need access to a GPU so will most likely be using Paperspace Gradient for the initial model creation.

The stack I'll most likely be using is & exploring is:

- Jupyter (local)
- Paperspace Gradient
- FastAPI
- Docker
- AWS (EC2, RDS)
- Streamlit