E-commerce Product Ranking using Text and Image Embeddings

Andrew Bregger, Janardhana swamy adapa, Nikita Duseja, Sai Eswar Epuri.

What is your research question? Clearly define the research problem/question

The research question that we are trying to solve is to build a product search engine by using both image, text and descriptive features for products in the dataset. We would evaluate and compare the results with only descriptive(baseline),text+descriptive, image+descriptive and descriptive+text+image feature embedding obtained from the dataset.

Why is this an interesting question to ask and why would we care about the answer to this question or a solution to the problem.

Ranking of query relevant results is one of the most important design criterion for search platforms and the chief evaluation metric for search relevance. In the realm of ecommerce search platforms with huge amount of visual and text data, and the current advancements in Deep Learning, analyzing and getting insights from high dimensional feature spaces is possible. Including the image data in building the search engine will be an interesting method to analyse the search ranking, as compared to the ranking obtained from just product descriptions and categorical descriptors. Using both images and text components in building Information retrieval components for ecommerce, successful results can open a lot of opportunities in other areas like recommender systems.

Has any existing research work tried to answer the same or a similar question, and if so, what is still unknown?

Some works have tried using both image data and description to predict labels for the products. In our approach we are considering a more comprehensive and unsupervised approach where we take the text descriptions and categorical descriptors of the product and get important labels using entity recognition for specific features for training the deep neural network and then combine both image features and text features embeddings to get the overall representation of the product. What we are trying to figure out and what is unknown is whether the combined representation is more useful than just the text or image based engine.

How do you plan to work out the answer to the question. (At the proposal stage, you are only expected to have a sketch of your methods.)

Dataset: Our dataset contains e-commerce product descriptors like image of product, text description, and reviews, and other attributes of the product.
Approach: we will extract specific entities and features from the text descriptions. Product image embeddings obtained from fine tuning the pretrained networks combined with the text

features, we aim to rank the products against the given query and compare our rankings with the rankings from common ecommerce platforms, mainly ebay and walmart.
We would evaluate and compare the results with only descriptive (baseline), text+descriptive, image+descriptive and descriptive+text+image feature embedding obtained from the dataset. We hope for an improvement as we embed visual embeddings into our ranking model.

How would you evaluate your solution. That is, how do you plan to demonstrate that your solution/answer is good or is reasonable.

We are going to evaluate using the Crowdflower search relevance dataset where there are around 32,000 products and for each query they have given top 25-50 results from different ecommerce sites such as ebay and walmart. The Crowdflower dataset contains a query, the resulting product, description of the product, median relevance of the results, and the variance of the relevance. The data set is already divided into training and testing sets where the testing set has the relevance scores withheld. We are going to assume that the queries in the dataset are the golden truth and evaluate against our model using metrics like NDCG.

rough timeline to show when you expect to finish what. List a couple of milestones.

Week 1 - extraction of product reviews from web, extracting features from text descriptors
Week 2 - ranking using only text and other descriptors, training CNN models for images
Week 3 - combining semantic information from both image and text and building the query engine
Week 4 - Analysis of different models - only text, only image, image + text

References:

https://www.kaggle.com/c/crowdflower-search-relevance/data
http://cbonnett.github.io/Insight.html
http://cs231n.stanford.edu/reports/2017/pdfs/105.pdf