Image Classification for Chatbots in E-commerce

Mahmoud Jobeel (*Computer Science and Engineering Technology*) and **Prof. Mohammed Salem** (*Digital Media Engineering*)

The German University in Cairo

amina.sakr@student.guc.edu.eg

Chatbots are now very popular when it comes to online shopping. These chatbots are usually used as a customer service to answer frequently asked questions from users which helps reduce waiting time. When the user is online to search for a product he wants to buy, he often searches using keywords such as Amazon and Jumia. Based on the data entered, the appropriate products are recommended to the user. Such search engines. If you enter information about a product you want to buy and you are not sure of it, the result will be inappropriate and boring. For example, if you see someone wearing a T-shirt and you like it a lot, going to search for it online will be tiring and waste of time to find something similar. Describing a product you want based on shape and texture is not easy. As a result, a capability that allows you to upload or take a picture of a desired item and recommends the same or similar items from the web site products dataset is required. Chatbots can be trained to provide this recommendation abality for a specific search query through image and text with the abaility to respond to the user questions

Literature Review

Paper [6] presented an idea of combining image recommendation and image recommendation decades ago. In this project, we use Amazon product dataset, which is used to build typical recommender system using collaborative filtering in [4] and [8]. In the field of image recommendation, [5] tends to recommend images using Tuned perceptual retrieval(PR), complementary nearest neighbor consensus (CNNC), Gaussian mixture models (GMM), Markov chain (MCL), and Texture agnostic retrieval (TAR) etc. CNNC, GMM, TAR, and PR are easy to train, but CNNC and GMM are hard to test while PR, GMM, and TAR are hard to generalize. Also, since data consists of images, the neural network should be a worth trying method. Paper [7] presented AlexNet model that can classify images into 1000 different categories. In adddition, paper [9] presented VGG neural network that classify images in ImageNet Challenge 2014. In our first part of project, we use both models to classify the categories of the products. However, both papers did not present a method for image recommendation. Although there are papers that studies image similarity such as [12] and [11], most of them are based on category similarity, i.e. products are regarded as similar if they are in the same category. However, products that come from the same category can still vary a lot. Thus, one reliable strategy is to first classify target image into a certain category and then recommend images from this classified category. In paper [13], they considered using neural network to calculate the similarities within category. However, the paper only consider ConvNet, DeepRanking etc. Since we have larger dataset, deeper convolutional neural network such as AlexNet and VGG should outperform naive ConvNets. The idea could be also found in [3] and [10]. Paper [1] is also focusing on learning similarity using CNN. However, it considers more on the case that multiproduct contained in a single image.

Literature Review

In our project, we assume that users are looking for a product and so image would only contains one product. Before we recommend, we need to answer What is the measurement of similarity. The most nature answer is either cosine similarity or L2 norm similarity. Another way to measure the similarity is by introducing semantic information

Methodology

The process to achieve product recommendation can be broken down in this few steps:

1) Transfer learning from ResNet-34 model(trained on ImageNet) to detect 37 classes in Fashion Product dataset using FastAI and Pytorch.
   Dataset source:
   https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small
2) Take the output of the second last fully connected layer from a trained ResNet-34 model to get embedding for all 38000 Fashion-37 images.
3) Use Locality Sensitive hashing to create LSH hashing for our image embedding which enables fast approximate nearest neighbour search.
4) Then given an image,we remove the background firstly using Modnet model, then we can convert it into image embedding using our trained model and search similar images using Approximate nearest neighbour on Fashion-37 dataset.

- Building text classification System:

1) We made a data file in JSON format so we used the json package to parse the JSON file into Python. The data file consists of tags and patterns collected by us. For example, we have a tag called price associated with patterns like “How much? ” , “What does it cost?”.

2)Preprocessing text data by breaking the whole text into small parts like words then applying cleaning such as removing stop words, noise removal, correcting spelling, Steammation and lemmatize each word and remove duplicate words from the list.

3) Using Bag of words algorithm as our text embedding technique.

4) building a multi classification deep neural network that has 3 layers..

Results

we calculated accuracy by simulating the way a user might judge the accuracy of our system. To do this, we ran through a randomly selected, statistically significant sample of images in our dataset (20% of dataset, consists of 7643) and examined the 5 best matches for each of these images. We have a score of 3 values (1 if out is the same category, 0.5 if similar to the category for example casual shoes and sports shoes, 0 others for example (T-shirt and shoes)). In other words, for query input we calculate the score out of 5 and finally get the average for all test data as the following equation:
recommended products/ total recommendations
Using this method, we were able to determine our recommendation engine and achieved 96.85% accuracy in suggesting similar products.
Also, I included the average time which each query takes and get 0.04s average time for each query.

Conclusion

Overall, 96.85% We are pleased with our progress thus far and look forward to further enhancing our work.
Future work:

1) Building a speech system along with text.
2) Building a mobile app which will be more helpful for such an application.
3) Implementing a recommendation system using text input.

References

[1] S. Bell and K. Bala. Learning visual similarity for product design with convolutional neural networks. ACM Transactions on Graphics (TOG), 34(4):98, 2015.

[2] T. Deselaers and V. Ferrari. Visual and semantic similarity in imagenet. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1777–1784. IEEE, 2011.

[3] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In Advances in Neural Information Processing Systems, pages 658–666, 2016.

[4] R. He and J. McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In Proceedings of the 25th International Conference 6 on World Wide Web, pages 507–517. I

Prepared for Thesis Poster Display Conference

11th -12th June 2022

Faculty of Management Technology
THESIS Poster Display Conference 2022