1. **Briefly descrive the problem you want to slove again, explaining what is the novelty (difference from prior work) in your project?**

* Problems: a solution that allows finding relevant products related to an image query even though their image reperesentation is not similar to the query
* The way to prove the solutions: re- constructuring model frame work to make best decision for images/video semantics and retrieval optimization with effectiveness.
* Detatil Tasks: When given data has the dataset of natural language form “retrieval keywords”, need to build the ranking system from candidate product lists.

(that images given by Alibaba paltform sellers. Which can score by top-ranked relationship with right images but there’s no support for products titles and explain texts. This research not for the text-image matching but for gauge fit-in model valuation between mutimodalities and language)

* Research novelty and differency:
  + There have been so many retrievals optimize matching with NLP query and images. But focus on multimodality features - especially international e-commerce platform – still have a lack of research. So, this thesis will be strongly related to multiple verification to cover the multimodality focusing. Also it need to understand the e-commerce and digital marketing knowledge.

So, as I will take a participate in KDD cup 2020, I will reasearch this topic which I’ve been interested in business filed.

* Baseline: Number 5. Has the answer detail. (move to the below number 5.)
* Key words: e-commerce, recommand, optimizer retrieval, natural language query mappaing with images, top ranked exposure product list tracking algorithm
* Background:

Most e-commerce and retail companies achieve their targets by leveraging the power of data and boosting sales by implementing search and recommender systems on their websites. With the above mentioned new upcoming display trends and huge increment of the online traffic volume, two fundamental challenges have to be solved. *First, we need to design a model framework to facilitate an effective semantic understanding, search, and retrieval of images and videos, so that we can recall and rank products with their appropriate multimodal contents.* This can help consumers make better decisions as well as delve more into the platforms with great potential. Second, machine learning systems are prone to exploitation if equipped with short-term goals, e.g., ctr, cvr or gmv, and the direct effects would cause severe Matthew's effects where only very small proportion of items from different sellers get exposed to online users. Thus, understanding whether the systems are fair is crucial and has deep influence on the development of the e-commerce platforms. To help tackle these challenges and further nourish the development in this field, we propose the following two competing tasks, recall for multi-modal entities and debiasing, as detailed below.

1. **Find at least 5 different related research work and wirte one paragraph for each paper on how they are related to your project use the following to search Google Scholar, ACM(ACM digital library), IEEE(IEEE explore)**
2. Understanding the way of multimodal analysis and frame work:

* [[paper]](https://www.sciencedirect.com/science/article/pii/S0268401218312751)[code]Multimodal data as a means to understand the leaning experience (Michail.N. et al., 2019)
* [[paper](file:///C:\Users\funny\Desktop\성균관대3학기_과제\data%20analysis%20application\homeworks\HW3\sciencedirect.com\science\article\abs\pii\S0306457316300206%3fvia=ihub)][code]A multimodal query expansion based on genetic programming for visually-oriented e-commerce applications(Patrícia C.SaraivaaJoão et al., 2016) – ranksSVM, random forest
* [[paper](https://www.sciencedirect.com/science/article/pii/S1084804519303078?via%3Dihub)][code]Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals (Nusrat J. Shoumy et al., 2020)
* [[paper](C://Users/funny/Desktop/%EC%84%B1%EA%B7%A0%EA%B4%80%EB%8C%803%ED%95%99%EA%B8%B0_%EA%B3%BC%EC%A0%9C/data%20analysis%20application/References/Flexible%20Fashion%20product%20retrieval%20using%20multimodality-base%20deep%20learning.pdf)][code] Flexible Fashion Product Retrival Using Multimodality-Based Depp Learning (Yeonsik jo et al., 2019)

1. Different approaches:

* [paper][code] A survey on recent image indexing and retrieval techniques for low-level feature extraction in CBIR systems. (In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Communication Technology, Chaziabad, india 13-14 Feb.2015)

1. Neural network approaches:

* [paper][code] Deep Fashion: Powering robust clothes recognition and retrieval with rich annotations (Liu,z.,et al., 2016)
* [paper][code]Sequence to sequence learning with neural networks. (Sutskever, I et al., 2014)
* [paper][code]Large-scale item categorization in e-commerce using deep learning and distributed computing. (Commun. Korean Inst. Inf. Sci. Eng. 2016, 34-32-38)

1. Top-performing approaches:

* [paper][code]Image question answering using convolutional neural network with dynamic parameter prediction. (Noh,H.,et al,. 2016 IEEE)
* [paper][code]LSTM: A search space odyssey (Greff,K at el., 2017)

1. Definitional approaches:

* TensorFlow. 2018. Available online: <https://www.tensorflow.org/community/contribute/docs> (accessed on 20 April 2020).
* Keras 2020. Available online: <https://keras.io/> (accessed on 20 April 2020)
* “Discounted cumulative gain” Wikipedia.org: <https://en.wikipedia.org/wiki/Discounted_cumulative_gain> (accessed on 23 April 2020)
* “Evaluation measures” (information retrievla): <https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval)> (accessed on 23 April 2020)

1. **Describe more details about your dataset**

**(of samples, labels, features etc)**

① Dataset versions.

There’re 2 versions of dataset bundles. Internationl version and china(domestic)version.

For research data, I’ll choose the international version. prohibit for only focusing on china maintain bias and I thought that Int’ data will be possesed whole customer interactions which can decide the some of important features related with multimodalities.

② Consisit of dataset.

1. Multimodal\_train\_sampleset.zip (82.53MB)
2. Multimodal\_labels.txt (8056MB)
3. Multimodal\_train.zip(26.19GB) – unzipped from that file, train.tsv can be accessed for training records. Before released the offical whole set of trainning data, sampled set containing 10k pair of data. Each ow in .tsv files represents a single sample pair. It has 9 columns separated by tables.

tabs:

1. product\_id: the index of the product
2. image\_h: the height of the product image
3. image\_w: the width of the product image
4. num\_boxes: the number of detected object bounding boxes for the image
5. boxes: a [num\_boxes, 4] shaped 2-D array specifying the location of each object bounding box in the image. For each row of this array, the 4 values specify the top/left/bottom/right coordinates of the corresponding box. This array is encoded into a base64 string, you can recover it into a numpy-array using np.frombuffer(base64.b64decode(THIS\_ENCODED\_BASE64STRING), dtype=np.float32).reshape(NUM\_BOXES, 4) in python
6. features: a [num\_boxes, 2048] shaped 2-D array specifying the 2048-dimension feature computed by the detector of each object bounding box in the image. This array is also encoded into a base64 string, you can recover it into a numpy-array using np.frombuffer(base64.b64decode(THIS\_ENCODED\_BASE64STRING), dtype=np.float32).reshape(NUM\_BOXES, 2048) in python
7. class\_labels: a [num\_boxes] shaped 1-D array specifying the category-id of each object. There are 33 object classification categories in this dataset. The correspondence between each category-id and its category name is given in multimodal\_labels.txt. The array is also encoded and can be recovered by np.frombuffer(base64.b64decode(THIS\_ENCODED\_BASE64STRING), dtype=np.int64).reshape(NUM\_BOXES)
8. query: a natural language query which matches the corresponding product
9. query\_id: the index of the query

1. Mulimodal\_validpics.zip(2.41GB)

* This files has the ground-truth product images can access the origianl image dataset.

1. Multimodal\_valid.zip(137.2MB)
2. Multimodal\_testA.zip(269.38MB) – testB set will be uploaded before 2weeks when competition end comes up. (finalized performance test will be testB set)
3. Multimodal\_submit\_example\_testA.zip(25.73KB)

③ Valid & testing set

Valid set : consist of 500 queries

Test set A/B : consist of 1K queries

Above sets will be prepare a candidate pool of around 30 products for ranking.

In this dataset each row representation candidates product – query fair and training set / valid set are considering as the same schema in 9 rows.

Valid set and test set’s query is respecitvely but “product id” is sharing same code system in valid/ test set.

Valid\_answer.json has real product base on evaluation evaluation queries. (it’s not sorted and ranked in order)

This sets format like this:

{

"query-id":

[

"ground-truth product-id 1",

"ground-truth product-id 2",

...,

"ground-truth product-id n"

]

}

④ submission format

Format need to be “sumission.csv” foramt and titles.

Each row has to be started with query-id and model searching with devide 5 top-tier product id. Csv file included the header and following format.

query-id,product1,product2,product3,product4,product5

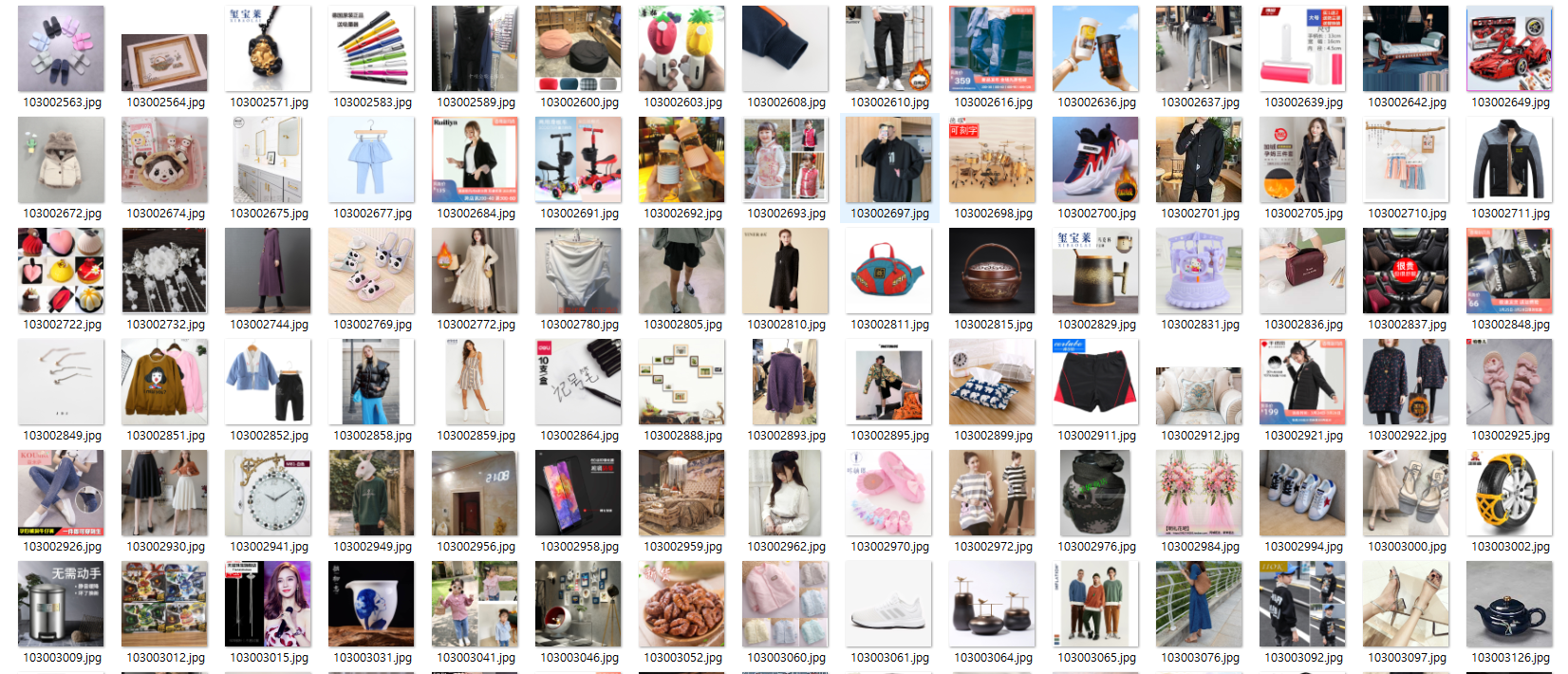
10001,10000001,10000003,10000007,10000002,10000010

10002,10005001,10000013,10000207,10000102,10000008

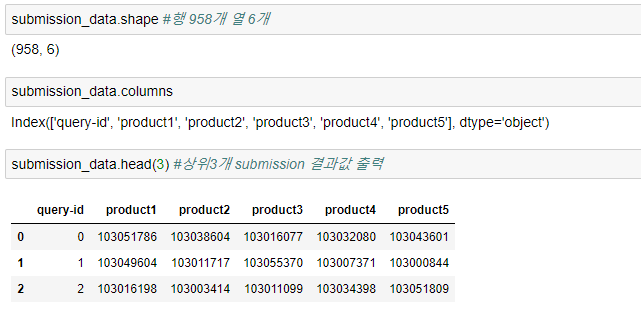
....

And also included the summation of report(suggested length 1-4 pages)

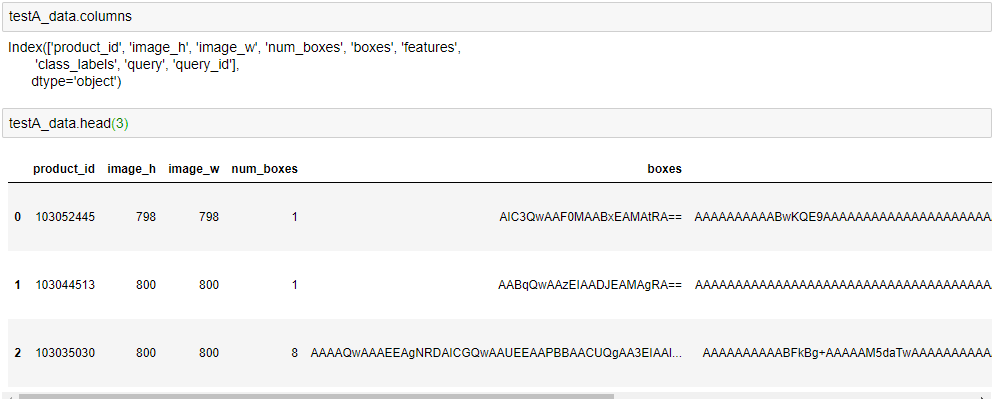
1. **Show three samples from your dataset and their detailed structure/fileds**



↑About 9K pictures given by Alibaba.com



↑Submission samples



↑testA\_dataset samples

1. **Provide the details of the baseline methods**

I didn’t decide for what method will be right fit-in this dataset,

But base on prior research, I could categorize as the retrieve optimizer, choosen following this way :

* Color based
  + RGB historgram
* Texture-based
  + Garbor filter
* Shape-based
  + Daisy
  + Edge histogram
  + HOG(histogram of gradient)
* Deep methods
  + VGG net
  + Residual net
* Incase feature mix : need to re-engineering extract the feature and ensemble
* Incase Dimension reduction : prior freature mix has dependency still have to considering issues.
* Etc : this dataset has .tsv (tab devided) depending on treate .tsv file need to conversion .tsv -> .csv file expandation.

1. **Provide the details of your evaluation metrics (e.g., Predictions, Recall, MAE?)**

First, I enforce with slove that tasks, I thought the best way for this is evaluating F1 precision and recall. So that I would like to use the way of F1 method for perfomance evaluation.



But after have some research, incase gauge performance especially in ranking quality and information retrieval parts, I found definitely (or almost..) used the nDCG(Discounted cumulative gain) algorithm. And also, the competition referer will determind by the nDCG@5(Disconted cumlutive gain) on testing set. So, final evaluation metrics will be nDCG@5.