Dialogue-based Image Retrieval

Intro

- Task Details
- Dataset Details
- Some examples of the data

Task Details

1. **Input:**The MSCOCO caption of the target image, followed by a dialogue which discusses the image.
2. Predict the correct image from a line up of 10 images.
3. Levels:
    a. Easy: Target image + nine other random images with no overlap of objects with the target
    b. Hard:
        i. At most five other images with same objects as the target image
        ii. Remaining images have same supercategory or intersection of object with target image

Evaluation Metrics

Quantitative

- Top 1 accuracy
- Top 5 accuracy

Qualitative

- Observations made on hard and easy datasets w.r.t various parameters you can think of.

Minimum Models

- Naive CBOW
- CBOW + basic NLP preprocessing
- RNN, (LSTM or GRU)

Dataset details

- There is only one dialog for each target image
  - Train: 40k dialogs
  - Validation: 5k dialogs
  - Test: 5k dialogs
- Dataset creation
- Dataset format and structure

Data Sources

- VisDial, https://visualdialog.org/data
  - Dataset created from the training split of the original dataset
  - The structure of the given dataset is different from the original i.e, data preprocessing is already done
- MS COCO, http://cocodataset.org/#home
  - Object data used to create the other nine images

Categories and Supercategories

Number of objects: 80(but the category index goes till 90)

Number of supercategory: 12

```
{'person': 0,
'animal': 3,
 'electronic': 9,
 'vehicle': 1,
 'furniture': 8,
 'outdoor': 2,
 'food': 7,
'accessory': 4,
 'kitchen': 6,
 'indoor': 11,
 'appliance': 10,
 'sports': 5}
```

Categories and Supercategories

```
{'id': 1, 'name': 'person', 'supercategory': 'person'},
 {'id': 2, 'name': 'bicycle', 'supercategory': 'vehicle'},
 {'id': 3, 'name': 'car', 'supercategory': 'vehicle'},
 {'id': 4, 'name': 'motorcycle', 'supercategory': 'vehicle'},
 {'id': 5, 'name': 'airplane', 'supercategory': 'vehicle'},
 {'id': 6, 'name': 'bus', 'supercategory': 'vehicle'},
 {'id': 7, 'name': 'train', 'supercategory': 'vehicle'},
 {'id': 8, 'name': 'truck', 'supercategory': 'vehicle'},
 {'id': 9, 'name': 'boat', 'supercategory': 'vehicle'},
 {'id': 10, 'name': 'traffic light', 'supercategory': 'outdoor'},
 {'id': 11, 'name': 'fire hydrant', 'supercategory': 'outdoor'},
 {'id': 13, 'name': 'stop sign', 'supercategory': 'outdoor'},
 {'id': 14, 'name': 'parking meter', 'supercategory': 'outdoor'},
 {'id': 15, 'name': 'bench', 'supercategory': 'outdoor'},
 {'id': 16, 'name': 'bird', 'supercategory': 'animal'},
 {'id': 17, 'name': 'cat', 'supercategory': 'animal'},
…..
 {'id': 41, 'name': 'skateboard', 'supercategory': 'sports'},
 {'id': 42, 'name': 'surfboard', 'supercategory': 'sports'},
 {'id': 43, 'name': 'tennis racket', 'supercategory': 'sports'},
 {'id': 44, 'name': 'bottle', 'supercategory': 'kitchen'},
 {'id': 46, 'name': 'wine glass', 'supercategory': 'kitchen'},
```

```
{'id': 58, 'name': 'hot dog', 'supercategory': 'food'},
 {'id': 59, 'name': 'pizza', 'supercategory': 'food'},
 {'id': 60, 'name': 'donut', 'supercategory': 'food'},
 {'id': 61, 'name': 'cake', 'supercategory': 'food'},
 {'id': 62, 'name': 'chair', 'supercategory': 'furniture'},
 {'id': 63, 'name': 'couch', 'supercategory': 'furniture'},
 {'id': 64, 'name': 'potted plant', 'supercategory': 'furniture'},
 {'id': 65, 'name': 'bed', 'supercategory': 'furniture'},
 {'id': 67, 'name': 'dining table', 'supercategory': 'furniture'},
 {'id': 70, 'name': 'toilet', 'supercategory': 'furniture'},
 {'id': 72, 'name': 'tv', 'supercategory': 'electronic'},
 {'id': 73, 'name': 'laptop', 'supercategory': 'electronic'},
 {'id': 74, 'name': 'mouse', 'supercategory': 'electronic'},
 {'id': 75, 'name': 'remote', 'supercategory': 'electronic'},
 {'id': 76, 'name': 'keyboard', 'supercategory': 'electronic'},
 {'id': 77, 'name': 'cell phone', 'supercategory': 'electronic'},
 {'id': 78, 'name': 'microwave', 'supercategory': 'appliance'},

….
 {'id': 87, 'name': 'scissors', 'supercategory': 'indoor'},
 {'id': 88, 'name': 'teddy bear', 'supercategory': 'indoor'},
 {'id': 89, 'name': 'hair drier', 'supercategory': 'indoor'},
 {'id': 90, 'name': 'toothbrush', 'supercategory': 'indoor'}
```

Dataset format and structure

Format : JSON

Files : data, metadata and dataset creation ipython notebook

Image Features: ResNet-152 features for all the images will be given

Structure: (description and structure for metadata and image features will be given later)

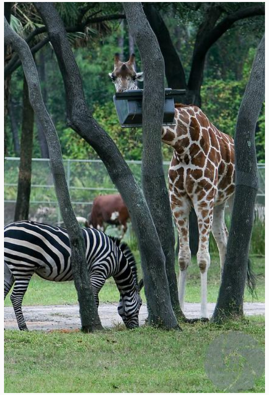
```
{ 0:
    { "dialog": [ [exchange 1], [exchange 2]........[exchange10]],
      "Img_list" : [ list of image ids],
       "target": index of target image in the image list,
        "target_img_id": 378466
    },

.....
}
```

Example: Easy

```
"4":
 {"dialog": [["is this a zoo ? yes"],
              ["how many giraffes are there ? 1"],
               ["how many zebras ? 1"],
               ["are people there ? no"],
               ["what did the giraffe eat ? not sure it is eating out of bin facing other way"],
               ["is it sunny ? yes"],
               ["how many trees are there ? 6"],
               ["what colors is the feeding bin ? black"],
               ["are there stripes on it ? no"], ["is the wind blowing ? not sure"]],
"caption": "a giraffe takes food from a feeding bin high up on a tree next to a zebra grazing on the grass",
"img_list": [35102, 386203, 323213, 379433, 461501, 259316, 411571, 176478, 332243, 553984],
"target": 8,
"target_img_id": 332243}
```

Images for previous Image List



PyTorch Review and Reference Tutorials

1. General Dataloader tutorial, http://pytorch.org/tutorials/beginner/data_loading_tutorial.html
2. PyTorch Tutorials, http://pytorch.org/tutorials/index.html
3. Advanced PyTorch tutorials, https://github.com/yunjey/pytorch-tutorial

Github Repo for data, https://github.com/AashishV/NLP1_IR