Yelp Restaurant Photo Classification

by Anna Rumyantseva

Abstract

In December 2015, Yelp proposed "Restraurant Photo Classification" challenge on kaggle.com website. The challenge implies assigning sets of labels (e.g. "good for lunch", "has alcohol", "takes reservations" etc) to different business ids based on the corresponding users uploaded photographs of restaurants. The training set for the project consists of manually labelled images and corresponding business ids. To approach the problem, transfer learning on pre-trained Convolutional Neural Networks (CNN) was implemented. In particular, pre-trained CNNs were used to extract feature vectors for each image. Subsequently, multiclass support vector machine classifier was applied. The implemented approach resulted in xx score.

Introduction

This capstone project aims to build a classification model that assigns multiple labels to photos of restaurants. The capstone project is based on a Kaggle Competition provided by Yelp company. Currently, restaurant labels are manually selected by Yelp users when they submit a review. Selecting the labels is optional, leaving some restaurants un- or only partially-categorized. But Yelp's users upload an enormous amount of photos every day alongside their written reviews. This data set of restaurants photos can be turned into multiple tags using an automatic classification model. Yelp is an American multinational corporation headquartered in San Francisco, California. It develops, hosts and markets Yelp.com and the Yelp mobile app, which publish crowd-sourced reviews about local businesses.

Image Classification problem is a computer vision task that involves assigning an input image a label from a set of categories. Specifically, based on the set of images with assigned labels, one needs to build a prediction model that assigns labels to novel input images. Image classification has a wide range of practical implementations. Yelp Kaggle competition provides a great opportunity to implement image classification techniques.

Convolutional Neural Networks (hereafter CNN) revolutionized the field of computer visions by significantly outperforming previous techniques used for image recognition and classification. CNNs transform the original image layer by layer from the original pixel values to the final class scores. In practice, it is not feasible to train an entire CNN from scratch since it is computationally expensive and suitable data sets for training are rare. However CNN can be used as a fixed feature extractor for the task of interest – transfer learning approach.

In this project, the transfer learning on a pre-trained CNN was implemented to assign multiple labels to different business ids based on corresponding images provided by Yelp users. Further description of the data set and methodology is given in the next section.

Data and Methods

Yelp dataset.

The data set for the project is Yelp users' uploaded images (.jpg file format) of restaurants and corresponding labels and business ids. The Data Files can be downloaded from site: (https://www.kaggle.com/c/yelp-restaurant-photo-classification/data).

List of the Data files:

- train_photos.tgz - photos of the training set (235841 images; 6.64 GB)

- test_photos.tgz - photos of the test set (474304 images; 6.71 GB)

- train_photo_to_biz_ids.csv - maps the photo id to business id

- test_photo_to_biz_ids.csv - maps the photo id to business id

- train.csv - maps the business ids to their corresponding labels.

There are 2000 business ids in the training data set and 10000 in the test one.

There are 9 labels for business IDs that can be assigned to photographs (Table 1).

Table 1. Description of the restaurant labels.

Label #	Description
0	good_for_lunch
1	good_for_dinner
2	takes_reservations
3	outdoor_seating
4	restaurant_is_expensive
5	has_alcohol
6	has_table_service
7	ambience_is_classy
8	good_for_kids

Pre-trained CNN model

To extract the feature vectors (also known as CNN codes) from the Yelp images, VGG ConvNet model developed by Karen Simonyan and Andrew Zisserman was implemented. The network was trained on the ImageNet data set (1.2 million images with 1000 categories) and became a runner-up for the Large Scale Visual Recognition Challenge 2014 (ILSVRC2014). The winner of the challenge was GoogLeNet. But, it was shown that VGG ConvNet features outperform those of GoogLeNet in multiple transfer learning tasks.

The images were processed through the pre-trained model using MatConvNet Matlab toolbox. MatConvNet is a MATLAB toolbox implementing CNNs for computer vision applications with many pre-trained CNNs for image classification are available in the toolbox. Fast VGG architecture (imagenet-vgg-f.mat 216 Mb; scheme of CNN is available here http://www.vlfeat.org/matconvnet/models/imagenet-vgg-f.svg) was downloaded for the feature extraction. Average image (available in net.meta.normalization.averageImage) was subtracted from all the images before running them through the model. The features vectors were taken as an outcome of the fully connected layer #7. The length of each vector is 4096.

Classification

Based on the obtained CNN codes the following steps were implemented for assigning labels to business ids:

1. Calculate mean CNN code vector for each business id in train and test data sets ignoring duplicates.

2. Train Multiclass Support Vector Machine classifier (SVM-multiclass) on the train set that consists of mean CNN codes for each business id and sets of labels listed in Table 1.

3. Make prediction on the mean CNN codes for unlabelled business ids.

Results

The Fig 1. represents a word cloud of image scores from VGG-Convnet model derived from Yelp images. The word cloud shows that the absolute majority of images scores are "food-related", with "plate" score being the most common.
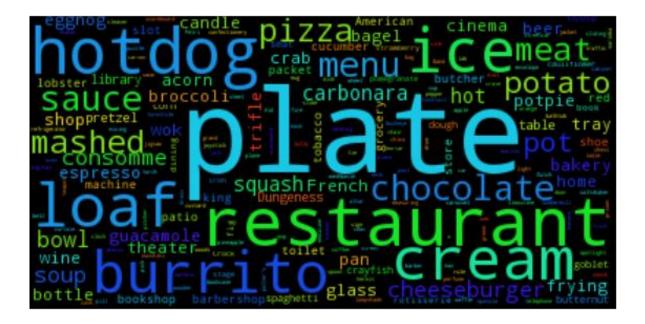

Figure 1. Images scores from VGG Convnet deried from Yelp images.