Multilabel Tagging System for Yelp Restaurants with User Uploaded Photos

Haomin Peng Boning Huang Yiqing Ding
{haomin, bnhuang, yiqingd}@stanford.edu

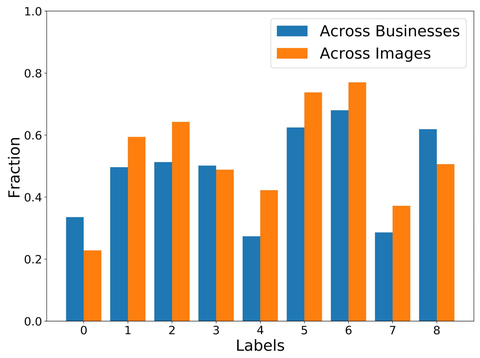
Motivation

Yelp is a multinational company that publishes crowed-sourced reviews of local businesses. Millions of people upload photos of businesses to Yelp everyday and thus provide valuable information to other users. Though these photos are typically tagged with simple information such as business names, deep level information, like whether the restaurant is good for lunch or dinner, usually requires manual labelling. This tedious process thus provides a great opportunity to apply machine learning techniques to real life.

Data Description & Preprocessing

- Description:

Kaggle provides a comprehensive dataset from Yelp. We are given photos of businesses and asked to predict the attributes of these businesses. Nine attributes are provided and each business can have more than one attribute. There are a total of 234841 photos belong to 2000 restaurants in training set, and each businesses can have more than one image. We are provided the labels and which picture belong to it in training set, and a test set with only photos and their belonging business id.



❖ Fig. (Left). Fraction of each label in the dataset; (Right). Sample images from the dataset. (a, b): labelled with “good_for_dinner”, “takes_reservations”, “alcohol” and “table_service”; (c, d): labelled with “takes_reservations”, “outdoor_seating”, “expensive”, “alcohol”, “table_service” and “classy_ambience”.

- Preprocessing:
  - Split given trainset: 80% for training and 20% for train validation
  - Resize images to 224 * 224

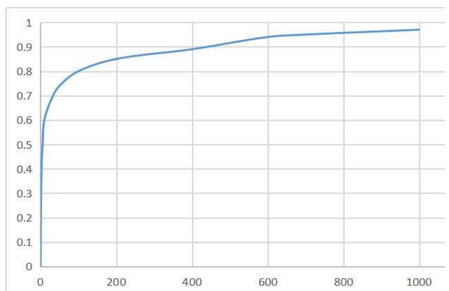


Fig. PCA analysis . The first 600 principle components contribute to 95% of variance of the original data.

Method and Model

Pipeline:

- Model 1: Perform classification on features extracted by PCA
- Model 2: Do transfer learning on resnet
- Model 3: Extract features with pretrained resnet model then build 9 xgboost classifiers on features

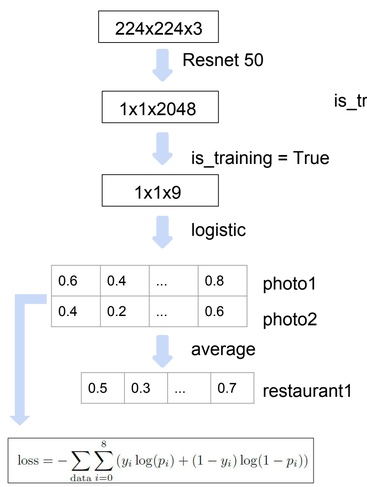


Fig. Model 2 pipeline

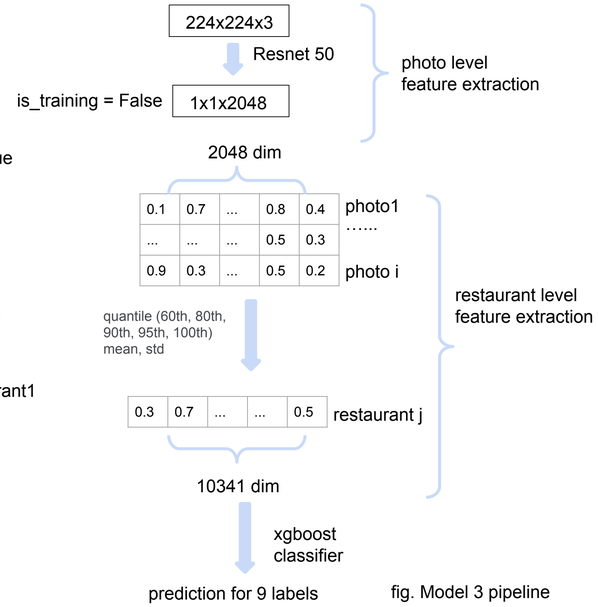


fig. Model 3 pipeline

Photo Level Feature Extraction:

- feed preprocessed photos into resnet, extrat the last average pooling layer, got features of dimension 2048 for each photo

Restaurant Level Feature Extraction:

- for photos of each business, calculate 60th, 80th, 90th, 95th, 100th percentiles, mean and std deviation of the 2048 photo-level features
- remove features with low std in training set, then we got 10341 features for each of the 2000 restaurants

Train classifiers:

- use One VS Rest (OVR) strategy to train 9 xgboost classifiers

Evaluation

	F1 on train set	F1 on train_val set
PCA+SVM	0.684	0.669
PCA+logistic regression	0.692	0.661
resnet50	0.762	0.758
resnet100	0.746	0.752
resnet50 + xgboost	0.829	0.802

- we evaluate the performance of our model with F1 score defined as:

$$F1 = \frac{2pr}{p+r}$$

Future works

- Extract photo features from more layers of CNN, not just the last pooling
- Design more clever loss function to weigh the multilabel loss