Ambience is classy or good for kids? : Yelp Restaurant Photo Classification Challenge

Jee Ian Tam

Computer Science Department, Stanford University, Stanford CA 94305

Objectives

- Predict labels corresponding to a business, given multiple images of the business taken by users.

- Metric - F1 Score on test set of 10,000 businesses :

$$\frac{2pr}{p+r} \qquad p : \text{precision} \qquad r : \text{recall}$$

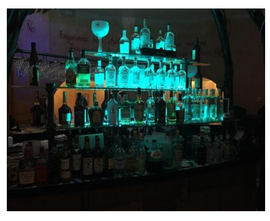
Rewards moderate performance on both over extreme performance on one at expense of another.

- Investigate various problem and algorithm transformations for multi-label classification.

Business Label Set

- Good for lunch
- Good for dinner
- Takes reservations
- Outdoor seating
- Is expensive
- Has alcohol
- Has Table Service
- Classy Ambience
- Good for kids

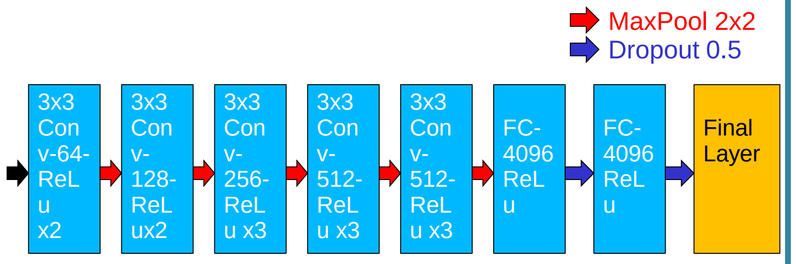
Samples from Training Set



- Takes Reservations
- Has Alcohol
- Has Table Service
- Good For Lunch
- Outdoor Seating
- Good For Kids
- Expensive
- Has Alcohol
- Classy Ambience

General Approach

- Start with pre-trained VGG-16, progressively fine-tune layers starting from last (highest) layer.
- ConvNet Architecture :



Simple Multi-Instance Learning

- Assign image labels from business labels
- Train on and predict labels from images
- Predicted (binary) business label is average of constituent predicted image labels, threshold at 0.5.

Problem Transformation 1 : Binary Relevance

- **Train 1 binary classifier for each label.**
- Equivalent to replacing final layer of VGG-like net with N output neurons having sigmoid activations, and setting loss as binary log-loss. N = number of labels.
- Advantages:
  - Simple interpretation and straightforward to implement
- Disadvantages:
  - Does not take into account correlations between labels

Problem Transformation 2 : Label Powerset

- **Classifier over every possible combination of labels.**
- Use softmax output layer followed by categorical cross-entropy loss
- Advantages :
  - Takes into account correlations between labels
  - Turns multi-label classification into multi-class classification problem
- Disadvantages :
  - Number of final layer neurons and weights exponential in the number of labels. Easier to overfit, sparsity can be an issue.

Algorithm Transformation: BP-MLL

- **Use loss that considers label correlations**
- Adaptation of backpropagation for multi-label learning

$$E = \sum_{i=1}^{m} E_i = \sum_{i=1}^{m} \frac{1}{|Y_i||\overline{Y}_i|} \sum_{(k,l)\in Y_i \times \overline{Y}_i} \exp(-(\hat{c}_k^i - \hat{c}_l^i)).$$

Output Activations; Labels in i-th training sample; Labels not in i-th sample

- Minimizing loss corresponds to maximizing the difference between activations of in-set labels and out-of-set labels.

- Use a variant of BP-MLL that allows for implicit learning of thresholds of each label while optimizing the network.
- Number of final layer neurons : 2 x number of labels.

- Advantages :
  - Considers correlations between labels, linear complexity
  - Fits well with ConvNets, thresholds learned during optimization.
- Disadvantages :
  - Trickier to implement, more pre-processing needed.

- Write loss function in Theano, pass as objective to Keras – No need to derive gradients for backprop!

Training

- For each method, we train on ~100,000 samples.
- Data augmentation :
  - Brightness, Shift, Rotation, Horizontal Flip, Crop
- Layers are progressively fine-tuned. Batch size = 20

Fine-Tuned Layers	Epochs	Initial Learning Rate
Last layer	2-3	1e-3
Top 16 (~40% of net)	4-5	1e-4
Top 32 (~80% of net)	5-6	1e-5

- Learning rate is divided by 10 when val. loss plateaus.
- L2 weight regularization parameter : 5e-4
- SGD with momentum 0.7. RMSProp used for 1st layer.

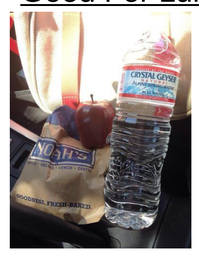
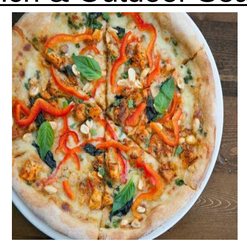
Results

Method	F-1 Score on Test Set
Binary Relevance	0.67
Label Powerset	0.62
BP-MLL	0.75

Benchmark with Color Features : 0.64
Current Leaderboard Standing : 36 / 170
Current Leaderboard Top Score : 0.83

Samples from Test Set Predictions

Good For Lunch & Outdoor Seating & Good For Kids


Is Expensive & Classy Ambience & Has Alcohol