Yelp Dataset

(Image Classification & Chatbot Based Recommendation)

Video URL - https://www.youtube.com/watch?v=nudb7iET7Q0

Devesh Kandpal
Soham Das Neogi

Final Project : INFO 7390 Advances In Data Science

Overview

Yelp.com is the go-to website for anybody looking for recommendations about local businesses. Over the period of time, it has become quite popular for restaurant recommendations as users find it easier to judge a local business by reading reviews about that business posted by other users. In addition to writing reviews and rating local businesses, user can react to reviews, plan events or start up a discussion forum. Yelp has about 135 million monthly customers and 95 million reviews.

Goals

The Yelp dataset is open sourced and it has enough information about each local business, reviews about those businesses , tips written by the user about a local business and photos uploaded by the user about the local business. This dataset can help us summarize useful information about most popular and unpopular restaurants classified by food genre and user reviews. For each user we can see his social connections and understand how much of an influencer can he be on other users. Lastly, we have an entire dataset comprising of images uploaded by different users with manually class tagging done on each image.

Our goal is handle the following use-cases:
1. Using image dataset comprising of labeled data, we would like to create a Deep Learning Model that is capable of recognizing objects and giving them an appropriate label. Our goal would be to start with predicting a single label correctly and then extend it towards dealing with multi-class classification.
2. By using reviews and tips dataset we would like to train a Deep Learning Model that would recommend restaurants to users based on the mood of the user. We would be developing a contextual chatbot for this purpose.

Use Cases

A. Local Businesses

Our predictive model can help us automatically tag images onto one (or maybe more) categories, thus avoiding human intervention.

B. Everyday User

Using the chatbot we would should be able to get restaurant recommendations by mapping user emotions against reviews of local businesses.

Data

1. **Yelp Dataset**: http://www.yelp.com/dataset
   We will work with the Yelp Dataset available online.

Process Outline

1. Data Preprocessing
- Data Cleaning, handling missing values
- Joining Business.json, Reviews.json, User.json, Tips.json and Image.json
2. Exploratory Data Analysis
3. Understanding Tensorflow / Tf-learn / Keras, for building a Deep learning Model that we would use for image classification.

4. Understanding how to build a Contextual Chatbot using Tensorflow / Tf-learn / Keras
5. Build an automation pipeline using Luigi for data ingestion and cleaning.
6. Deploy the Model on AWS or Google Cloud Computing Platform
7. Build a web application to demonstrate the image classification and chatbot implementation.

Milestones

Timeframe	Delivery
Day 1-3	Data Preprocessing and Exploratory Data Analysis
Day 4-9	Model Building, Training (Image Classification & Chatbot)
Day 10 - 12	Deployment of models on cloud and build web application
Day 13-14	System integration and documentation

Deployment / Infrastructure Details

1) Language: Python, Java
2) Pipeline: Luigi
3) Cloud Tools/Platforms: AWS / Google Cloud Platform
5) Tools for Analysis: Tableau / Power BI

Reference Sources

https://www.yelp.com/dataset/documentation/json
https://chatbotsmagazine.com/contextual-chat-bots-with-tensorflow-4391749d0077