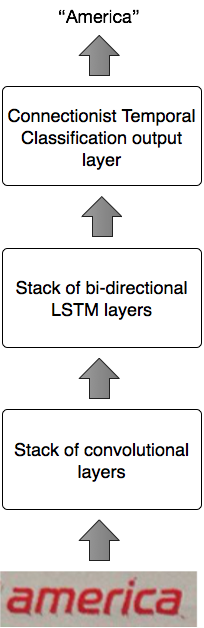
OCR

Optical Character Recognition (OCR) is a software that converts printed or handwriten text using images or other mediums into vector digitized form. OCR is a piece of software required for machines because standalone machines are not intelligent enough to understand the text from images. OCR is a combination of various Computer Science disciplines including NLP, Image Processing, Pattern Classification, etc. Based on the type of input, the OCR systems can be categorized as handwriting recognition and machine printed character recognition. The former is relatively simpler problem because characters are usually of uniform dimensions, and the positions of characters on the page can be predicted. Handwriting character recognition is a very tough job due to different writing style of user as well as different pen movements by the user for the same character. Our system will be using OCR for printed text that is recognizing text from images of price labels.

OCR is a composite activity comprises different phases. First, an external source captures an image of text labels. Then various type of preprocessing is performed to improve the quality of image. Different types of filters such as averaging, min and max filters can be applied. Alternatively, different morphological operations such as erosion, dilation, opening and closing can be performed. The characters in the image are separated such that they can be passed to recognition engine. Among the simplest techniques are connected component analysis and projection profiles can be used. The segmented characters are then processes to extract different features. Based on these features, the characters are recognized. Character classification maps the features of segmented image to different categories or classes. There are different types of character classification techniques. Structural classification techniques are based on features extracted from the structure of image and uses different decision rules to classify characters. After classification, the results are not 100% correct, especially for complex languages. Post processing techniques can be performed to improve the accuracy of OCR systems. These techniques utilize natural language processing, geometric and linguistic context to correct errors in OCR results. This includes spell checking and dictionaries.

Dropbox has published a blog post explaining their OCR technology achieved using Computer Vision and Deep Learning. They have created this system specifically for mobile handheld devices. The first task was to label a picture data. This will be used for training the models later on. The labeling was done on user collected data by human intervention. The labeling was performed by specific teams at DropTurk and MTurk. This labeling is considered as a ground truth for individual images. The next job is to detect words from an image. This problem is divided into 2 parts. That is dividing the images into images containing one line per image and using these images to get actual text using Deep Neural Networks. Images are fed to Convolutional Neural Network. The output of CNN is then fed to Bidirectional Long short Term Memory Which is used for speech recognition and consequently, making words from range of characters.



The main problem of using deep learning is that the neural nets typically require a huge amount of dataset for achieving good accuracy. Getting original data from user is difficult. Dropbox used another Deep Learning model called Generative Adversarial Models which is well suited for generating realistic data. All this process gave more accuracy for commercial use on one single word. To get the system working for entire document, the paragraphs of words are broken down into separate lines for input to the neural nets. This is achieved by Maximally Stable Extremal Regions using OpenCV’s implementation. The models were first developed in Torch but were imported to Tensorflow.

Location

Our application needs a large amount of database of stores across USA. This makes the information complex and hard to filter. We are considering an addition of location based filtering. This will allow the user to get the store data that is nearer to them faster, and efficiently. This will also allow us to implement pub-sub system for users favorites and shopping lists.

Here we discuss some research done in this area. The first paper implements a recommendation system based on location. Recommendation system can consider user’s location for recommending the nearest service (ATM, restaurants, pubs, tourist sights and social events). The MOPSI project implements various location-based services and applications such as mobile search engines, data collection, user tracking and route recording. It has applications integrated both on web and in mobile phones.

Aim in MOPSI is to recommend interesting places in user’s surrounding. The system gives personalized recommendations by combining various paradigms of recommendation systems. It uses location as a preselection criteria. The data that is present far away from the user is irrelevant. After that the sorted list is created from the pre-selection. Services are scored using contextual information about search history, location, and explicit rating.