Recipe Recommender System Using Image Recognition of Food Ingredients

Mona Mishra, Yifan Gong

Abstract

In this paper, we propose a recipe recommendation system employing image recognition of food ingredients, specifically vegetables. Currently the system is a web application that performs image recognition on the uploaded images and recommends recipes that contains the recognized ingredients. We built a convolutional neural network model for image recognition to identify five categories of food ingredients and achieved, 62.9% accuracy rate. The recommendation system uses the labels of the identified images to display a list of recipes that contains most of the identified ingredients.

Introduction

Websites or apps that have user profiles, use a recommender system that is based on the interests and the preferences of their users, such as Amazon.com, Netflix, YouTube, etc. These websites take note of their users current choices and compare them to their dataset to present a list of recommendations. The goal is to keep users interested by providing them relevant information in order to motivate them to come back to the website or app. There are several cooking apps or websites available today, that are used to find recipes based on some keyword, like name of the food ingredient or type of cuisine, etc. These apps are mindful of the needs and interests of their users, but they fail at identifying their user's constraints, i.e., limited number of food ingredients. In such cases, users find themselves shopping for ingredients or they decide to substitute the missing ingredient with something else. To help users avoid such adjustments, image recognition can be employed to identify food ingredients that are already available at their disposal and recommend them recipes based on those ingredients.

The main objective of the proposed system is to assist users to decide what they can cook with the available resources. We intend a user use our system not only at home while cooking but also during grocery shopping. By pointing the camera at the food ingredients, users will can immediately build a plan or have an idea of what they will be cooking that week, based on our recommendations.

Related Works

Image recognition is a classical problem in computer vision which identifies objects into one of finite sets of classes. A lot of research work has been done on image recognition. The ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2015) runs since 2010. It uses hundreds of categories and millions of images to do classification and detection tasks. Microsoft Computer Vision API as a well-known commercial service can describe and generate tags for an image.

There are many applications are based on object recognition, like Camfind and Memrise that classifies the object in an image. Camfind uses camera to capture an image and describes it using some of the essential features like color, shape, etc. It helps users to find products that are similar to the captured image. CamFind is created for people who want to learn more about an item, and perhaps shop for it online, without having to type in a search. Although it identifies fruits and vegetables, it is not trained to classify meat into categories like steak, chicken, etc. Memrise also performs image recognition but it is a language learning app that translates the label into the language that users is interested to learn.

In the food category, Researchers from MITs Computer Science and Artificial Intelligence Laboratory (CSAIL) developed an app called pic2Recipe (Salvador et al. 2017) that uses the photo of a prepared food and compares it to the images in its dataset to predict its ingredients and suggest similar recipes. They have developed 'Recipe 1M' dataset with the help of AllRecipes and Food.com to train a neural network to find patterns and make connections between the food images and the corresponding ingredients and recipes. On the other hand, there is Calorie Mama that provides information on calories based on the food images. It also recognizes food categories like fruits, vegetables, meat, etc. All of these systems are either based on food category recognition or recognition of generic objects that are not specifically food ingredients, while we focus specifically on food ingredients image recognition to provide recommendations.

Dubey and Jalal (Dubey and Jalal 2012) propose an approach to do fruits and vegetables classification. The major works are detecting segmentation of objects, extracting features and training with SVM model. In our work, we not only classify fruits and vegetables, but also add meat categories. After the classification task, we will also recommend recipes based on the recognition results.

Regarding the work on food ingredients recognition that provides recipe recommendations, the paper on ’Real-time Mobile Recipe Recommendation System Using Food Ingredient Recognition’ (Maruyama, Kawano, and Yanai 2012) proposes a method to recommend recipes based on the conditions of ingredients on amounts, nutrition and prices which is recognized from a short video. In our current work, we do not use the information on the amount of ingredient and its prices. We increase the number of categories of ingredients and use images replace video to do object classification. Then, we use the available food ingredients to present them a list of recipes.

In their paper, Machine Learning Based Food Recipe Recommendation System (Vivek, Manju, and Vijay 2018), Manju, Vivek and Vijaya talk about two approaches for recommending recipes, namely, item based approach and user based approach. They found that the user based approach performed better than the item based approach on the Allrecipes dataset that contained large number of interactions between users and items.

In the paper, Machine Learning Algorithms for Recommender System a comparative analysis (Sahu, Nautiyal, and Prasad), Sahu, Nautiyal and Prasad have exploited 5 different algorithms, namely, content-based filtering, collaborative filtering, hybrid content collaborative based filtering, k-means clustering and Nave Bayes classifier to understand their strengths and identify which of them gives the best precision and accuracy.

Gong (Gong 2011) talks about the problems associated with todays recommendation approaches, like data sparsity and scalability issues. She tries to resolve the scalability issues with item-based collaborative filtering by establishing relationship among items, but it worked poorly when the data was sparse. Finally, she employed an algorithm based on item semantic similarity and item rating similarity.

Our system is motivated by the approaches used in these paper but currently we present our recommendations by retrieving recipes from the dataset that contains the large number of identified ingredients. The recommendations also take into account that the recipes contain less number of ingredients that the user might not have.

Proposed System Prototype

A mentioned before, our system aims to search for recipes at home or during shopping at grocery stores. In this section, we describe the flow of how to use the proposed system taking the photos of the food ingredients. The photos will be detected by the system to recommend the recipes based on the objects recognized. The system interface is built using the flask framework.

Step1: UI displays a button to upload images and user successfully uploads images of food ingredients

Step 2: The recommendations are displayed based on the recipes that contained the items in the images that was searched.

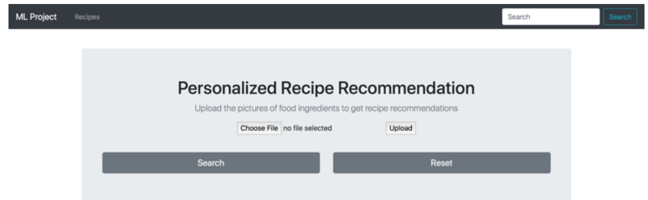

Figure 1: Step 1-1

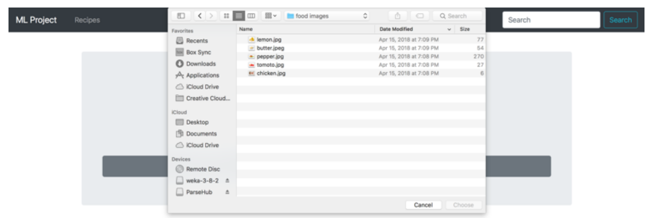

Figure 2: Step 1-2

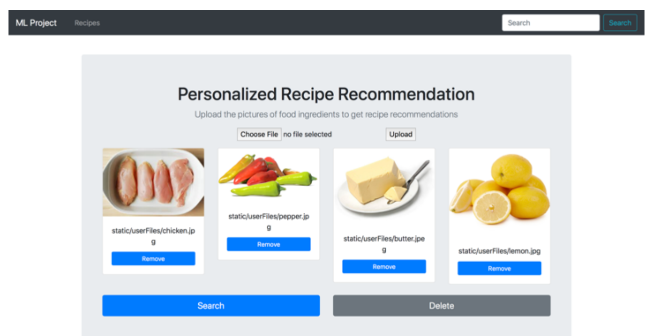

Figure 3: Step 1-3

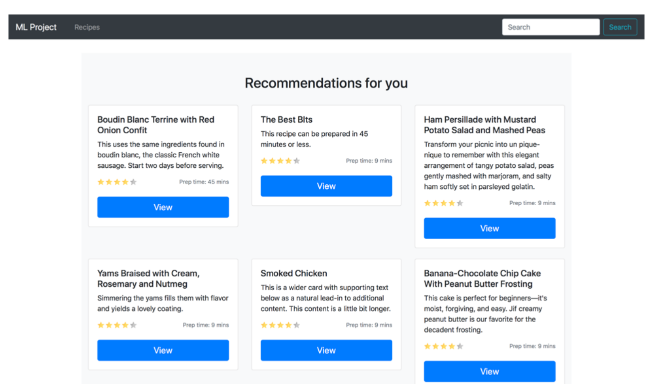

Figure 4: Step 2

Methods

Workflow

Our system aims to recommend recipe using the ingredients information provided by users. It allows users to upload ingredients images and recognize the object in the image. After recognizing ingredient images, users can get a list of recipes by a recommendation system. See work process in figure5

Image Recognition

In image recognition task, we build a convolutional neural network (CNN), train a few thousand images for multiple categories, and predict what is the object in the image from users upload.

We use Keras deep learning library in python to build our CNN model. Our CNN model contains a convolution layer to get the feature map, a pooling layer to reduce the size of the image, a flatten layer converting a two dimensional array to a one dimensional vector, a hidden layer with 128 fully-connected notes and an output layer. Before training with dataset, we use the pre-processing library in Keras to process images to avoid overfitting. Training dataset contains 25000 images. Images belong to the same category are kept in one folder. We defined ten epochs in training process to fit images to neural network. The predict result is printed as the category name of ingredient.

Recommendation System

In order to get a list of recipes, we send the names of recognized food ingredients as a keyword to the dataset and obtain the results containing a list of recipes. The ingredients in the dataset was in the following format: '1 1/2 pounds of chicken thighs' or '2 cups water'. With the help of pandas library functions and the text analysis API from Microsoft Azure, we cleaned the dataset so that we could perform successful search operations on it. After cleaning, the ingredient list contained the names of actual ingredient only, not their quantities. After performing search operation, the result consisted of a list of recipes that had at least one or more recognized ingredient in them. The search results are then ranked according to the number of keywords (ingredient) present in their list of ingredients.

Experimental Design

We experimented different categories combination from the dataset we built and fitted training data to three different training models to test the result and evaluate the performance of our system. The three training algorithms are support vector machine, a 5-layer convolutional neural network and ResNet50 provided by Keras library.

Dataset

We picked up twenty two fruits and vegetables categories from ImageNet (Deng et al. 2009). ImageNet provides image urls, original images and category names which are collected from Amazon Mechanical Turk (AMT). As there is not any image based dataset including different types of raw meat. We collected 1000 images per category for three kinds of meat. Therefore, our dataset contains twenty five different food ingredients including fruits, vegetables, meats and others.(see in Table 1). To provide recommendation, we used the Epicurious dataset from Kaggle that contained 20,000 recipes listed by ingredients, recipe rating and nutritional information.

Table 1: Food ingredients in the dataset

types	ingredients
meat	steak, salmon, chicken
vegetable	broccoli, cabbage, carrot, celery, corn, cucumber, eggplant, green bean, green pepper, olive, onion, potato, spinach, tomato
fruit	apple, avocado, banana, lemon
others	bread, cheese, mushroom, egg

Support Vector Machine(SVM)

In this experiment, we used 22000 images only from ImageNet to train the model. We extracted features from those images and use the Support Vector Machine (SVM) to classify them. We first trained SVM with single feature including histogram of oriented gradients (HOG), local binary patterns (LBP) and color features (RGB). Then, we trained it with a combination of HOG features and LBP features. We carried out evaluation of object classification performance with 10-fold cross validation.

SVM is a supervised machine learning algorithm and it is widely used in image classification. HOG is used to detect the edges and corners information in the image. HOG converts an image to a feature vector and gets the distribution of directions of gradients. Gradients of an image are useful because the magnitude of gradients is large around edges and corners. LBP is to get pattern and texture information. LBP is a texture descriptor which convert image to grayscale and select a neighborhood of size R surrounding the center pixel. The value is calculated for the center pixel. RGB values contain the color feature of an image.

Convolutional Neural Network

We built a simple convolutional neural network for image classification task with five layers. We started training the model with twenty two categories from ImageNet, but we realized that the training as well as the test accuracy was very low. So we reduced the number of categories to five and added some images from google to improve our dataset. Also, we tested the number of epoch from 1 to 10 to find the optimal number of epoch. We used the loss function and performance metrics to evaluate the result.

Keras provides a pre-trained model named ResNet50 (He et al. 2016). It is a 50-layers neural network trained with 1000 categories from ImageNet and won the first prize on the ILSVRC 2015 classification task. We used this model to evaluate our model.

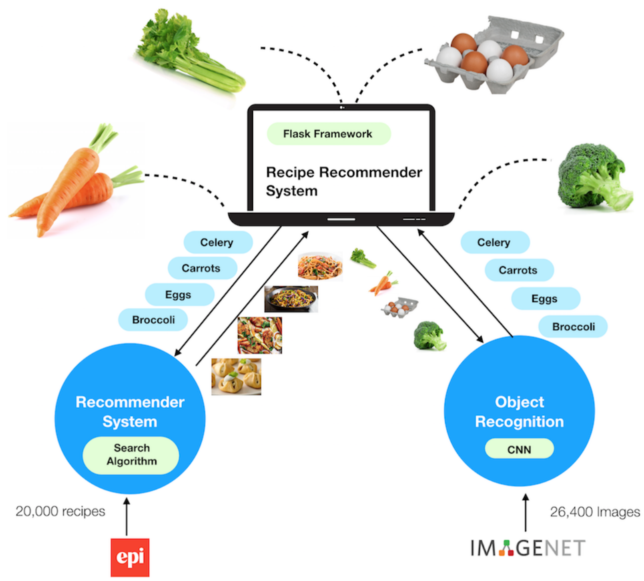

Figure 5: System Flow

Experiment Result

We evaluated the results of SVM model with training it using different image features. The accuracy of SVM with HOG feature is 0.014. The accuracy of using single feature of LBP or RGB is zero. The combination of HOG and LBP features also get a zero accuacy.

ResNet50 performs very well and it has an error rate of 3.57. But we can not modify the 1000 categories or change the category number. As our system is used to predict ingredient only. This model is not suitable to the system.

Our 5-layer convolutional network seemed to do very well with dataset that contained images that was a combination of images collected from google as well as imagenet. We trained the our CNN model with five categories. The accuracy and the loss plot between training data and validation data is depicted in figure 6 and 7. The CNN model did quite well compared to the SVM model that we created earlier, with the training accuracy of 99% and the validation accuracy of 62.9%. However, the model looks like it is overfitting since the training accuracy is quite high compared to the validation. We also used the CNN model to train all 25 categories. The validation accuracy declined to 23%. We kept changing the number of categories and found that adding categories causes a decline of the result's accuracy.

Currently, the recommendations are based solely on food ingredient. The results contains the list of recipes that has at least of the searched ingredient. In order to personalize the recommendations, we need to consider user's search history so that we can learn more about the recipes that they like and refine the search result accordingly. For this purpose, we are looking into association rules so that we can determine the items that are frequently searched or viewed by the user.

In future, we plan to improve the system in terms providing more accurate recommendations using the combination of nutrition and user reviews. For this purpose, we will

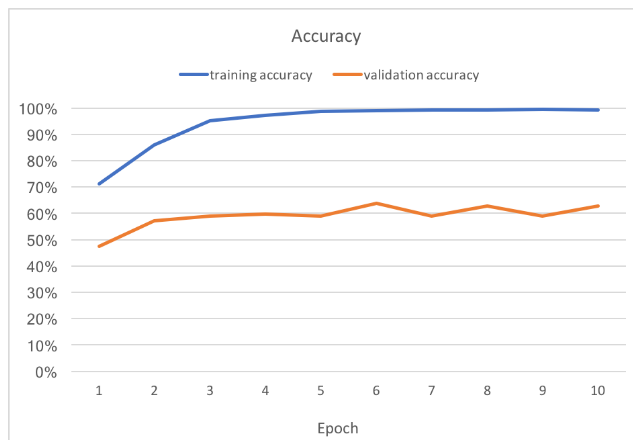

Figure 6: Accuracy comparison

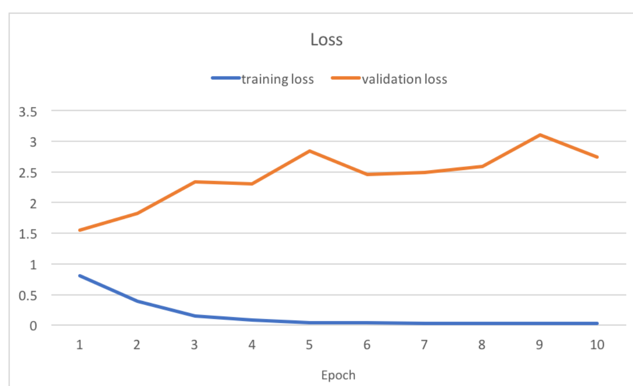

Figure 7: Loss comparison

need additional dataset of user reviews that would help us build user profiles using collaborative filtering techniques by identifying similar users based on their ratings and reviews. Our current system is based on a web application that forces users to upload images on the web platform in order to get recommendations. As our next step, we would like to improve upon the user interface and create a mobile application to provide ease of use and perform usability testing in an efficient manner. Besides, we hope to achieve multiple objects recognition in one image by using region convolusional neural network so that users can take a picture of all the ingredients they have to get recommend recipes.

Conclusion

Through this project, we learnt to create good datasets and build different models to improve upon the accuracy for object recognition. How we build the dataset affects the model greatly and hence it extremely important to clean the data before using it. We also found that as we increased the number of categories of the food ingredients, the accuracy dropped significantly. Hence, as we realized that in order to increase the number of categories, it is important to increase the number of images for all other categories to achieve best possible result.

References

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 248–255. IEEE.

Dubey, S. R., and Jalal, A. 2012. Robust approach for fruit and vegetable classification. *Procedia Engineering* 38:3449–3453.

Gong, S. 2011. A personalized recommendation algorithm on integration of item semantic similarity and item rating similarity. *JCP* 6(5):1047–1054.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Maruyama, T.; Kawano, Y.; and Yanai, K. 2012. Real-time mobile recipe recommendation system using food ingredient recognition. In *Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices*, 27–34. ACM.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115(3):211–252.

Sahu, S. P.; Nautiyal, A.; and Prasad, M. Machine learning algorithms for recommender system-a comparative analysis.

Salvador, A.; Hynes, N.; Aytar, Y.; Marin, J.; Ofli, F.; Weber, I.; and Torralba, A. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Vivek, M.; Manju, N.; and Vijay, M. 2018. Machine learning based food recipe recommendation system. In *Proceedings of International Conference on Cognition and Recognition*, 11–19. Springer.