BBM406 INTRODUCTION OF MACHINE LEARNING PROJECT REPORT

WHAT THE FOOD - ESTIMATING THE PRICE OF FAST-FOODS

Berk Sevgi
21328414

Ahmet Ayan
21327661

ABSTRACT

The basis of this project is to estimate price of whichever fast food by using Machine Learning methodologies. The main idea of that project which has been constructed with Convolutional Neural Network concept that is one of the most popular concept in Machine Learning community is to interpret a given image as basic classification problems with using pre-learned Neural Networks and make price assesment with respect to result of these interpretations.

1 INTRODUCTION

In sociology, it's explicitly known that there is a proportional correlation between economy and people's life style. People have been planning long or short-term plans regarding their life with respect to their economical power. In today's world with limited resources, people can spend money to satisfy their needs and prospering. According to a research conducted by Turkish Statistical Instutue (TK), foods have share in consumer price index with 23,68% rate. This information shows that peole have to spend money in important share like 23,68% of their salary to satisfy their food requirement.

Nutrition is one of the most important requirement of people to spend their lives. Every living has to eat enough and balanced nutrition to continue healthy life. Rapidly wasting resources have been the greatest obstacle to reaching the required food for the survival of the living. As a result of rapidly wasting resources, mankind is giving more money to food day by day. As a consequence, food is a high-priority planning topic for mankind in daily life.

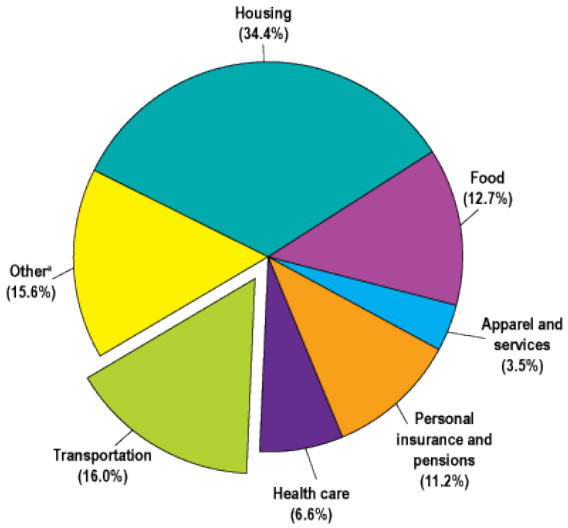


Figure 1: Average Household Expenditures by Major Spending Category: 2010 (Source : United States Department of Transportation)

Today's technology, like in many areas, is interested in solving the problems of ideal food with cheapest price as possible. In digital world, there are several applications providing alternative solutions about this problem and most of these applications use the machine learning concept.

Machine studies computer algorithms for learning to do stuff effectively and efficiently. The primary goal of machine learning research is to develop general purpose algorithms of practical value. These algorithms has been implementing under purpose of using in every problems.

Our goal is to constitute a Machine Learning structure that estimates price of a meal with respect to current food prices of real world by using its photo. A person who curious about the price of a meal will be able to learn estimated cost with using that program.

2 RELATED WORK

We've searched for related works for a long time. However, we could not find another project with the same aim as our project. We can divide the projects into two categories in general terms; - Projects that estimating using food - Projects that estimate the prices of other things These projects have short explanations in this title.

1.1- Projects that Estimating Using Food These projects process food photographs like ours and then estimate their properties like calories. Their approach inspired us, but it did not help much. Because the dataset they use does not usually work for us.

1.1.1- Calorie Estimation From Fast Food Images The project consists of two steps: identifying food from an image, and convert the food identification into a calorie estimation. They performed food image classification using SVM and deep learning algorithms. Different features such as LBP, HOG, and CNN are explored and compared. For the calorie estimation step, they created a calorie map for the image classification labels.

1.1.2- Recognition and Classification of Fast Food Images They aimed to utilize learnt machine learning algorithms to do fast food recognition. Their goal is to find a computational efficient algorithm with high accuracy. They used Bag of SIFT, color histogram and a combination of them as the features. k-NN and SVM method (with and without kernel) are used to classify fast food images to eight classes.

1.2- Projects that Estimate the Prices of Other Things The purpose of these projects is the same as ours, but the areas they work are different. There is no estimating the price of the food in these projects that estimate the price.

1.2.1- Machine Learning for a London Housing Price Prediction Mobile Application In this project, they presented a mobile application that can generate predictions for future housing prices in London. In order to select a prediction method, various regression methods were explored and compared. Gaussian Processes (GP) for regression was chosen as our model due to its flexible and probabilistic approach to learning and model selection. To handle the large dataset of past property transactions in London, they exploited the spatial structure of the dataset to distribute computations to smaller independent local models. Overall predictions are obtained by recombining predictions from local models. By training the model and generating predictions on the server-side of the application, they were able to offload computationally intensive tasks and focus on generating visualizations on the client-side. Their results demonstrate that our approach to the problem has been largely successful, and is able to produce predictions that are competitive to other housing price prediction models.

1.2.2- Predicting the Price of Used Cars using Machine Learning Techniques The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbors, nave bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy. All the four methods provided comparable performance. In the future, we intend to use more sophisticated algorithms to make the predictions.

3 DATASET

We have preferred to choose PFID : Pittsburgh Fast Food Image Dataset which was published in 2010. PFID : Pittsburgh Fast Food Image Dataset contains 4,545 images of fast food, 606 stereo pairs, 303 360 degree videos. We picked 1960 fast food photos which were captured in laboratory environment for the learning part. We have used 75% of all images (1470 images) for training operation in this dataset, which contains more than 100 products from 13 fast food companies, and the remaining 25% (490 images) for validation

Figure 2: Contents of Pittsburgh Fast-Food Image Dataset

4 USED MODELS

CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks are computational models in Machine Learning which has been designed for implementation of learning models, producing new informations and discovering inspired by human neural activities.

Convolutional neural networks are an important class of learnable representations applicable, among others, to numerous computer vision problems. A convolutional neural network structure contains one or more pattern which called as 'layer'. The convolutional layer is the core building block of a convolutional network that does most of the computational heavy lifting.

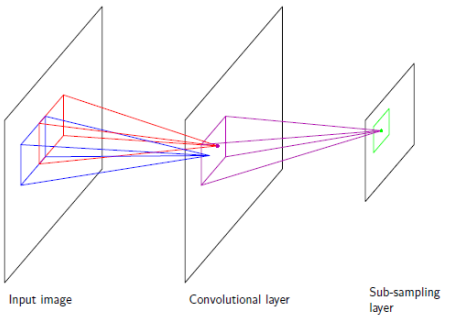


Figure 3: A simple convolutional neural network illustration.

Convolutional networks look at features of data desired to learn, make generalizations about them, collect information, and then decide on those examples using information they have learned before.

It is possible to represent neural network mechanism as a function demonstration.

$$f(x) = f_L(f_2(f_1(x; w_1); w_2)), w_L). \tag{1}$$

In this representation, f notions show results of each layers and w represent its weights.

ACTIVATION FUNCTION

Convolutional networks work on non-linear features. It's needed to convert linear dimension by using activation functions. There are several activation functions to convert non-linear features such as sigmoid function (σ), $tanh$ function or ReLU function.

Our network is based on Rectified Linear Unit (ReLU) function. The reason of usage this function is not saturating on positive feature values and it's available to compute derivation efficiently.

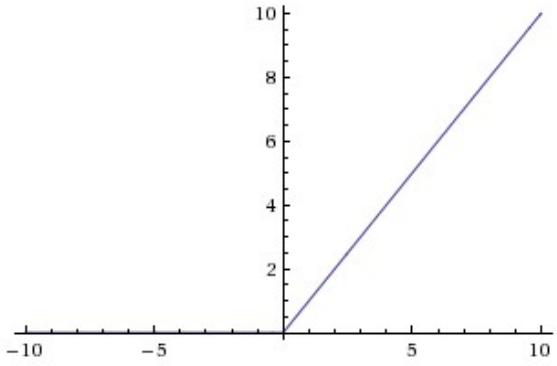
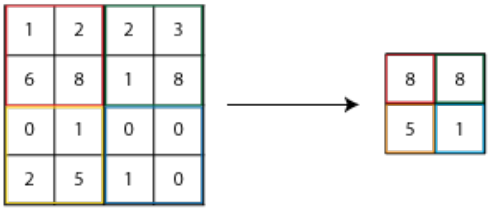


Figure 4: Rectified Liner Unit

POOLING

Pooling is a concept in deep learning visual object recognition that goes hand-in-hand with convolution. Pooling operation reduces the spatial size of the representation to reduce the amount of parameters and computation in the network.

Common pooling operations are max-pooling and sum-pooling. Max pooling decreases the dimension of input data simply by taking only the maximum input from a fixed region of convolutional layer.



$$y_{ijk} = max\{y_{ijk} : ii < i + p, jj < j + p\}$$

Figure 5: Max-pooling illustration and formulization are shown above

SOFTMAX CLASSIFIER

Softmax Classifier is a generalization of logistic regression that can be used for multi-class classification (under the assumption that the classes are mutually exclusive).

$$P(y = j|z^{(i)}) = \phi_{softmax}(z^{(i)}) = \frac{e^{z^{(i)}}}{\sum_{j=0}^{k} e^{z_k^{(i)}}} \tag{2}$$

$$L_i = -logP(y = j|z_{(i)})$$

Figure 6: Formulation of Softmax Classifier

5 APPROACHMENT

The data set consisting of 130x97 images was numbered to show training data 1 and verification data 2. In consequence of enumeration, training data was specified as amount of 75% of dataset and the rest was determined as validation data.

The training data set is sent for learning to the neural network consisting of 4 convolutional layers, 1 pooling layer and 1 activation layer. Each image in training dataset is converted size of 32x32 first. After translation, the resized images are converted to grayscale intensive images by using the $rgb2gray()$ function.

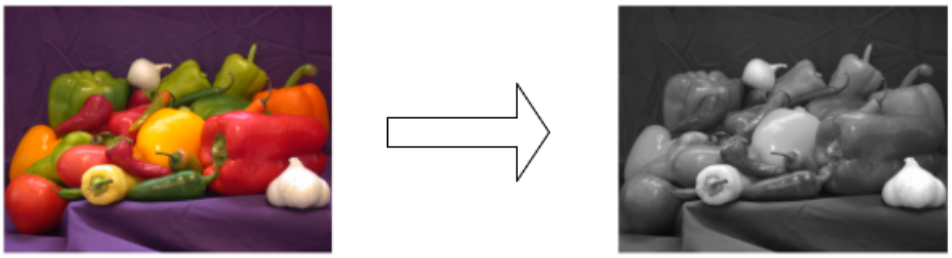
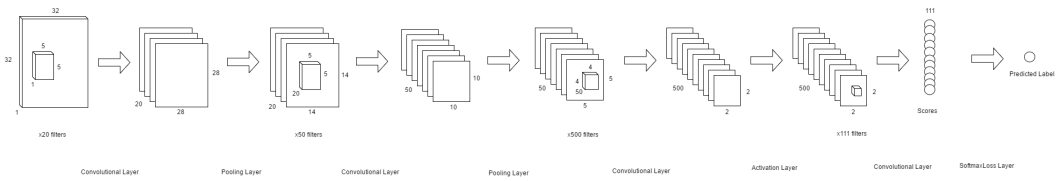


Figure 7: Transformation of RGB image to grayscale

After preprocessing of training images, learning process is beginning. The neural network which is consisting of 4 convolutional layer, 1 pooling layer and 1 activation layer is illustrated below.

In the first stage, 20 single channel convolution filters with 5x5 size create a 28x28 sub-sampling layer (pattern) with 20 channels from one training image. These newly created subsampling layers are processed in 2x2 pooling layers to create 14x14 sized patterns with 20 channels.Patterns with 14x14 size are processed in 5x5 convolution layer to create 10x10 sized patterns, after processing, newly generated patterns are compressed with 2x2 sized pooling layer and 5x5 sized new patterns generated. New patterns are processed in 4x4 convolution layers to create new patterns in 2x2 dimensions. Final patterns are compressed by 111 pooling layers with 2x2 sized and then 111 scalar



values are generated. These scalar values represents score values of training image corresponding to each label. The score values are solved by the softmax classifier to label the training picture as the label with the highest score.

Convolutional layer provides computational effectiveness by learning in data up to a certain batch size.Each batched data is re-learned by a certain number of iterations (epochs). The number of batches and epochs in the project are 50 and 30, respectively.

Since we think that the price estimate for food is collateral with prediction of food type, we decide to designed the program to actually predict food types.

6 Experimental Results

What is Top-N error?

Top-N error is the fraction of test samples x_i where the correct label y_i does not appear in the top NN predicted results of the model when results are sorted in decreasing order of confidence, or $P(yi|xi)$.

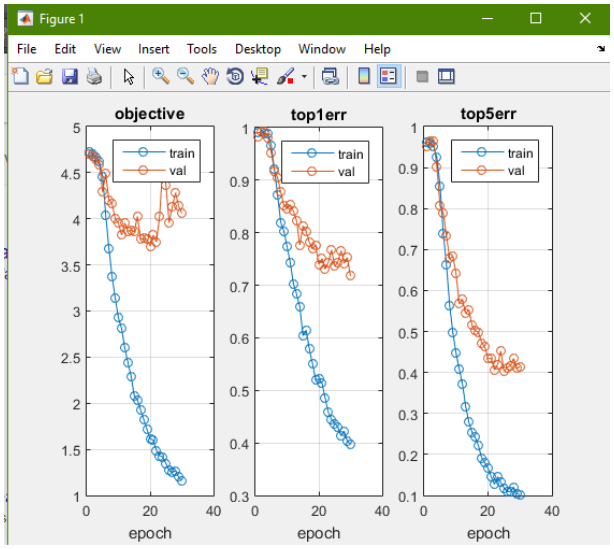


Figure 8: Top-1 and Top-5 error rates.

Top-5 error, also known as rank-5 error is simply an instantiation of Top-N error metric with (N=5).

As seen in figure above, our convolutional layer has almost %60 accuracy rate (in Top-5 scale) over the validation data.

7 References

1. ¡U.S. Department of Transportation, Bureau of Transportation Statics¿.

2. ¡CS231n: Convolutional Neural Networks for Visual Recognition/Stanford University¿.
3. ¡IEEE Xplore, PFID : Pittsburg fast-food image dataset¿.
4. ¡VGG Convolutional Neural Networks Practical by Andrea Vedaldi and Andrew Zisserman¿.
5. ¡What is the Role of the Activation Function in a Neural Network? by Sebastian Raschka, Michigan University¿.
6. ¡Sudeep Raja, A Derivation of Backpropagation in Matrix Form¿.
7. BISHOP, CHRISTOPHER 2006 *Pattern Recognition and Machine Learning*, Cambridge U.K.