ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Inverse Cooking Recipe Generation from Food Images and Cuisine Classification

Ruchi k. oza[1], Aanal S. Raval[1], Aakanksha V. Jain[1]

M. E Student, Department of Computer Engineering, Silver Oak College of Engineering and Technology, Ahmedabad, Gujarat, India [1]

ABSTRACT: In this paper, We introduce Inverse cooking system and cuisine classification. People enjoy food photography because they appreciate food. Behind each meal there is story described in a complex recipe and unfortunately by looking at food image we do not have its preparation process and from where this food belongs to. Therefore, in this paper we introduce image to recipe retrieval and cuisine classification. For this, we are using recipe1M+ dataset which is a new large-scale, structured corpus of over one million cooking recipes and 13 million food images. As the largest publicly available collection of recipe data, Recipe1M+ affords the ability to train high-capacity models on aligned, multimodal data. Neural joint embedding of image and recipe is using for retrieval sequence of recipe instruction, recipe name and ingredient from food images .we use generated ingredients for cuisine classification task for classification of cuisine This will help people to know the recipe information with its cuisine just from image.

KEYWORDS*:* feature extraction, joint neural embedding, classification

I. INTRODUCTION

Social People enjoy food photography because they appreciate food. Behind each meal there is a story described in a complex recipe and, unfortunately, by simply looking at a food image we do not have access to its preparation process. Therefore inverse cooking system that recreates cooking recipes from given food images. It predicts ingredients as sets by means of a novel architecture, modelling their dependencies without imposing any order, and then generates cooking instructions by attending to both image and its inferred ingredients simultaneously.
This system will help to the people who wants to know the recipe and cuisine of food from image. Also in the restaurant andchef for exploring more food dishes.

The inverse cooking system generates cooking instructions, title, and ingredients form the food images. We are using neural joint embedding .In embedding, the paired (recipe and image) data in order to learn a common embedding space. Embedding make it easier to do machine learning on large inputs like sparse vectors representing words.
Cuisine classification is classify cuisine from food ingredients generated by Inverse cooking system. It will classify the food by its types. There are many kind of cuisine like Indian cuisine, Italian cuisine, Mexican cuisine, French food, Thai food, Chinese food etc. it will help people to know the type of cuisine by from taking the picture of the food images. It is very useful for the people who don't know any knowledge about the food.

For inverse cooking system we address data limitations by introducing the large scale Recipe1M+[1] dataset which contain one million structured cooking recipe and their images for generating recipe from images.
For, cuisine classification task we use ingredients dataset which is already generated by inverse cooking system. We pre-process and train that data for classification task. In dataset, we gave ID to the ingredients and train dataset. System will able to predict the cuisine by using Classification algorithm.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

This system will able to predict the ingredients, instruction, name of food and type of cuisine. This system will help foodie people to know the recipe of food from food image.

II. LITERATURE SURVEY

For inverse cooking system, we have referreda methodology proposed to neural joint embedding model. Here, they utilize the paired (recipe and image) data in order to learn a common embedding space .they discuss recipe and image representations, and then, They describe our neural joint embedding model that builds upon recipe and image representations. For dataset, Recipe1M+ dataset is referred, a new large-scale, structured corpus of over one million cooking recipes and 13 million food images. As the largest publicly available collection of recipe data, Recipe1M+ affords the ability to train high-capacity models on aligned, multimodal data. Using these data, train a neural network to learn a joint embedding of recipes and images that yields impressive results on an image-recipe retrieval task. Moreover, they demonstrate that regularization via the addition of a highlevel classification objective both improves retrieval performance to rival that of humans and enables semantic vector arithmetic [1]

[2]This paper proposes a unique method of obtaining the compressed embedding of cooking instructions of a recipe image using cross model training of CNN, LSTM and Bi-Directional LSTM. The major challenge in this is variable length of instructions, number of instructions per recipe and multiple food items present in a food image. This model successfully meets these challenges through transfer learning and multi-level error propagations across different neural networks by achieving condensed embedding of cooking instruction which have high similarity with original instructions. They have specifically experimented on Indian cuisine data (Food image, Ingredients, Cooking Instruction and contextual information) scraped from the web. The proposed model can be significantly useful for information retrieval system and it can also be effectively utilized in automatic recipe recommendations.

For cuisine classification task we referrer[10],this paper proposed method for cuisine classification by doing data-preprocessing task levenshtein distance , TF-IDF and classification task . SVM is used in this paper for classification task.

III. PROPOSED METHODOLOGY

A. Dataset

We are using Recipe1M+[1] data set for inverse cooking system to predict the recipe from food images.Recipe1M+ is dataset contain one million structured cooking recipe and their images.it is already trained dataset. Figure 1. is a work flow of inverse cooking system and cuisine classification.

B. Feature extraction

We are using joint neural embedding [1]for recipe retrieval task. For feature extraction task we tried SE-RESNET-50, SE-RESNET-101 and SE-RESNET-50[5] is a combination of squeeze and excitation network and RESNET.
Se-resnet-50 have 50 layers of resnet, se-resnet-101 have 101 layers of resnet and se-resnet-101 have 101 layers of resnet convolution network.

C. Ingredients extraction

Ingredients extraction task we have used LSTM withword2vec embedding.
Representation of ingredient first step is word2vec embedding. Word2Vec is a shallow, two-layer neural networks

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

which is trained to reconstruct linguistic contexts of words. It takes as its input a large corpus of words and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the vector space.

Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located in close proximity to one another in the space. Word2Vec is a particularly computationally-efficient predictive model for learning word embedding from raw text .It comes in two flavors, the Continuous Bag-of-Words (CBOW) model and the Skip-Gram model. Algorithmically, these models are similar.

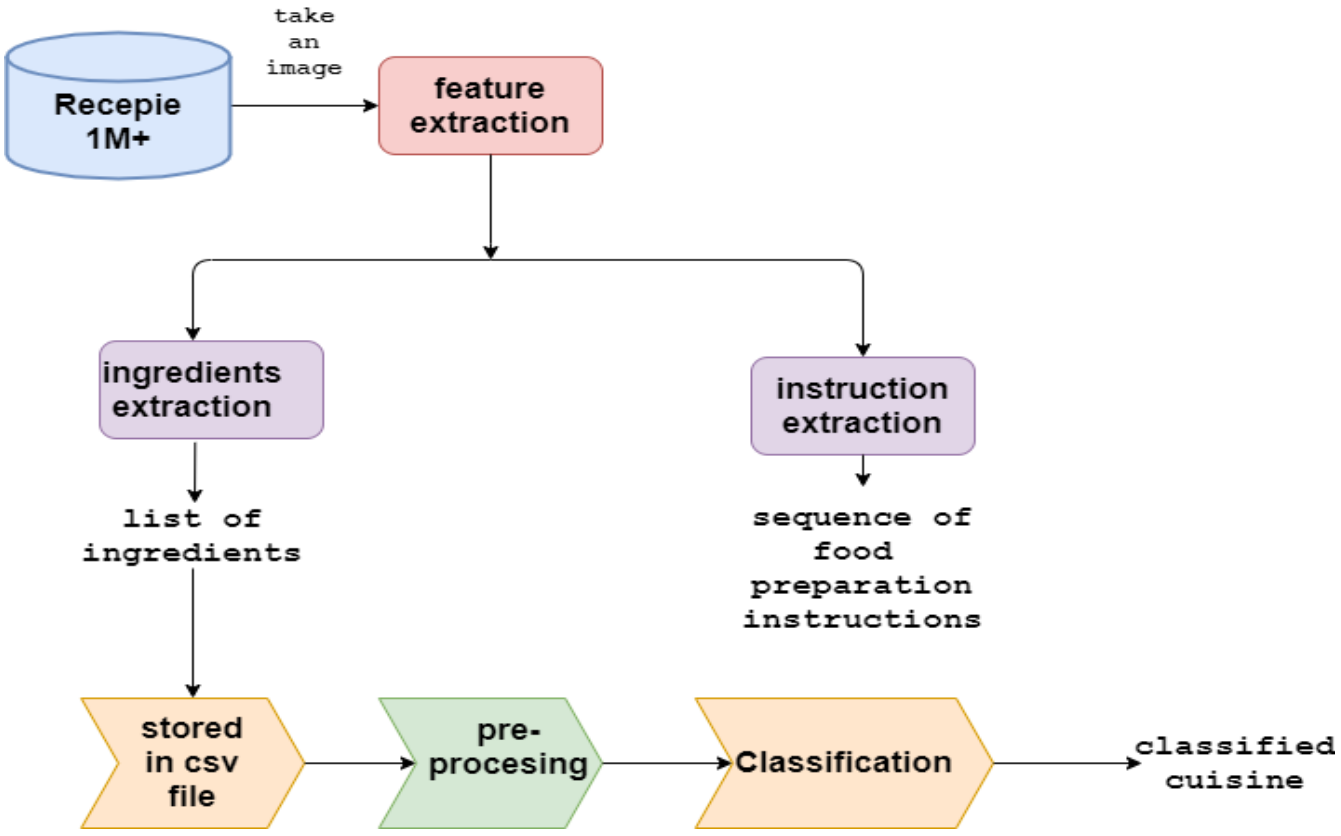

Figure 1.
Work flow of inverse cooking system and cuisine classification

I. Continuous Bag-of-Words (CBOW)

CBOW predicts target words (e.g. 'mat') from the surrounding context words ('the cat sits on the'). Statistically, it has the effect that CBOW smoothes over a lot of the distributional information (by treating an entire context as one observation). For the most part, this turns out to be a useful thing for smaller datasets.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

ii. Skip-Gram

Skip-gram predicts surrounding context words from the target words (inverse of CBOW). Statistically, skip-gram treats each context-target pair as a new observation, and this tends to do better when we have larger datasets.
Then second stage is LSTM. Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. There are two types of LSTM first is Bi-directional LSTM and second is encoder-decoder.

i. Bi-directional LSTM

The basic idea of bidirectional recurrent neural nets is to present each training sequence forwards and backwards to two separate recurrent nets, both of which are connected to the same output layer. This means that for every point in a given sequence, the BRNN has complete, sequential information about all points before and after it. Also, because the net is free to use as much or as little of this context as necessary, there is no need to find a (task-dependent) time-window or target delay size.

ii. Encoder-decoder LSTM

The idea is to use one LSTM to read the input sequence, one time step at a time, to obtain large fixed-dimensional vector representation, and then to use another LSTM to extract the output sequence from that vector. The second LSTM is essentially a recurrent neural network language model except that it is conditioned on the input sequence.

D. Instruction extraction

For instruction extraction LSTM is used. LSTM networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. There are two types of LSTM first is Bi-directional LSTM and second is encoder-decoder.

E. Cuisine classification

Cuisine classification task, we uses ingredients generated by inverse cooking system. We train that data for classification.

i. Pre-processing

First step is data pre-processing .we split data into train and test set, than TFIDF is performed.

ii. TF_IDF

TF-IDF (term frequency–inverse document frequency), is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. In text mining it is often used as a weighting factor. The paper discusses in detail about the TF-IDF. In their approach they will examine both the reduced and non-reduced feature vectors provided from the recipes collection. In addition, they use TF-IDF to create a more representative feature vectors for better classification.Term frequency is the original coarse count of term in document given by:

$$tf_{ij} = f / (\sum_i f_{ij})$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Inverse document frequency is a measure of how much information the word provides, i.e., if it's common or rare across all documents and is given by:

$$Idf_I = \log (N / df_i)$$

So, tf-idf is:

$$tfidf(i,j,I)=tf(i,j).idf(I)$$

iii. Classification

For classification, we tried SVM[3], XG-BOOST[6], CATBOOST[9] AND LIGHTGBM[8].
support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework.
CatBoost is a machine learning algorithm that uses gradient boosting on decision trees. It is available as an open source library.
Light GBM is a gradient boosting framework that uses tree based learning algorithm.

IV. EXPERIMENTAL RESULTS

In inverse cooking system, feature extraction task we tried resnet-50,se-resnet-50 and se-resnet-101.And ingredients extraction task we use word2vec embedding and bi-directional LSTM and instruction retrieval task we used two stage LSTM. We generate a sample dataset and perform different feature extraction algorithm.That contain ingrs.pkl ,instr.pkl and demo images file, From results , we get se-resnet-101 with higher accuracy than resnet-50 and se-resnet-50.

Table 1 shows the result of resnet-50, se-resnet-5- and se-resnet-101.

Model name	F1	Recall	Precision	Accuracy
Resnet-50	49.08	75.47	77.13	55%
Se-resnet-50	51.20	78.00	79.33	75%
Se-resnet-101	**85.259**	**84.92**	**85.600**	**85.20%**

Table 1.

Table 2. shows the accuracy of classification cuisine. We used ingredients which is generated by invers cooking system.we perform pre-processing, TF-IDF and classification. for classification SVM, XGBOOST, CATBOOST, light GBM is tested. Here are it's accuracy:

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Model name	Accuracy
SVM	72.12%
XG-BOOST	74.40%
Light GBM	**78.66%**
CAT-BOOST	76.68%

Table 2

Hence, as per results we have used SE-RESNET-101 with 85.20% accuracy for feature extraction task and classification task we used Light-GBM with 78.66%.

V. CONCLUSION

In inverse cooking system se-resnet-101 performs better than resnet-50 and se-resnet-50.In terms of accuracy Se-resnet-101 gives 85.20% accuracy. where resnet-50 and se-resnet-50 gives 55% and 75% accuracy respectively.

For cuisine classification task light gbmperforms better results in term of accuracy. Light gbm gives 78.66% accuracy. Where svm, xg-boost and cat-boost gives 72.12% , 74.40%, 76.68% accuracy respectively.This system will able to predict the recipe of given food image and it's cuisine.

Further work includes consideration of data from other food images from social media. Besides, links and videos also can be considered in dataset.

REFERENCES

[1] Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images Javier Mar´ın1 *, Aritro Biswas1 *, Ferda Ofli2 , Nicholas Hynes1 , Amaia Salvador3 , Yusuf Aytar1 , Ingmar Weber2 , Antonio Torralba1 1Massachusetts Institute of Technology 2Qatar Computing Research Institute, HBKU 3Universitat PolitecnicadeCatalunya{abiswas,nhynes}@mit.edu,{jmarin,yusuf,torralba}@csail.mit .eduamaia.salvador@upc.edu, {fofli,iweber}@hbku.edu.qa IEEE 09 july 2019 DOI: I 10.1109/TPAMI.2019.2927476

[2] Food Image to Cooking Instructions Conversion Through Compressed Embeddings Using Deep Learning MadhuKumariTajinder Singh Computer Science and Engineering Department Computer Science Engineering Department National Institute of Technology Chandigarh University Hamirpur, H.P., INDIA Panjab, INDIA madhu.jaglan@gmail.com nith2kl4@gmail.com 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)

[3] Analysis of classification models based on cuisine prediction using machine learning shobhnajayaramn ,tanupriya Choudhry , Praveen kumar IEEE

[4] Cuisine Prediction based on Ingredients using Tree Boosting Algorithms R. M. Rahul Venkatesh Kumar*, M. Anand Kumar and K. P. Soman Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, AmritaVishwaVidyapeetham, Amrita University, Coimbatore - 641112, Tamil Nadu, India; m_anandkumar@cb.amrita.edu, kp_soman@amrita.edu Indian Journal of Science and Technology, Vol 9(45), DOI: 10.17485/ijst/2016/v9i45/106484, December 2016

[5] Squeeze-and-Excitation Networks by Jie Hu1, Li Shen2, Gang Sun(Department of Engineering Science, University of Oxford), 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[6] XGBoost: A Scalable Tree Boosting System by Tianqi Chen (University of Washington tqchen@cs.washington.edu), Carlos Guestrin (University of Washington guestrin@cs.washington.edu), IEEE, August 13-17, 2016, San Francisco, CA, USA.

[7] Deep Residual Learning for Image Recognition Kaiming He Xiangyu Zhang ShaoqingRenJian Sun Microsoft Research {kahe, vxiangz, v-shren, jiansun}@microsoft.com

[8]GuolinKe, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma1 , Qiwei Ye, Tie-Yan Liu, "*LightGBM: A Highly Efficient Gradient Boosting Decision Tree",* 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA

[9]LiudmilaProkhorenkova, GlebGusev, AleksandrVorobev, Anna VeronikaDorogush, AndreyGulin, "*CatBoost: unbiased boosting with categorical feature",* 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

[10]Cuisine classification using recipe's ingredients S. Kalajdziski1, G. Radevski2, I.Ivanoska3, K. Trivodaliev4 and B. Risteska Stojkoska5 Ss.IEEE 2018