Determination of the belonging of a dish to the kitchen according to the ingredients

Irina Vankina, Alexey Komlev, Vasily Koreshkov

January 2022

Abstract

The paper describes an algorithm that implements the definition of the cuisine to which a given dish belongs, based on the ingredients used in cooking. The problem is solved using feed-forward neural networks, as well as using convolutional neural networks. You can see the results of the work and the source code at the link `https://github.com/heinwol/meal-type-from-ingredients`.

1 Introduction

Natural language processing(NLP) tasks are gaining more and more popularity nowadays. The authors of this work investigated the possibility of using NLP methods in matters related to cooking. Such tasks are relevant, since the introduction of machine learning technologies in this area will save time and money when searching for answers to various queries. In particular, the authors set the following task: to determine the cuisine which the dish belongs based on the ingredients in the recipe. Previously, the authors solved a similar problem – determining the category of a dish (breakfasts, drinks, pastries and desserts, etc.) according to the given ingredients. Many situations can be cited that confirm the usefulness of such tasks.

The complexity of the implementation lies in the fact that it is quite difficult to find a dataset suitable for such tasks. There are quite a few English-language datasets on the Internet. They have a rather limited set of fields; for the presented task, existing datasets need to be supplemented. Even less data is contained in datasets in Russian. In this regard, the authors decided to compile their own Russian-language dataset, which would be enough to solve problems related to cooking.

1.1 Team

Irina Vankina dataset processing, comparison of used algorithms, prepared this document.

Alexey Komlev the common concept of the task, dataset collecting, code development

Vasily Koreshkov dataset preprocessing, code development

2 Related Work

The problem related to predicting the type of cuisine by ingredients was previously considered in [1]. The dataset associated with the recipes of various dishes was collected, for example, by the authors of [2]. However, this dataset is in English. We also found a dataset with recipes in Russian [3], but it does not contain the fields "cuisine" and "category", respectively, it is impossible to solve our initial task with its help. Based on the ideas outlined in the publications above, we implemented the classification task for the dataset that we collected in Russian.

In our work, two models are considered. The first model is a fully connected neural network for classification by cuisines/categories of dishes. The second is the creation of a convolutional neural network to solve this problem. This approach is described in detail in the article [4].

3 Model Description

One of the main goals of this work was to collect a high-quality dataset in Russian, with the help of which it would be possible to solve the problem of classification in relation to dishes. A detailed description of the resulting dataset is presented in Section 4.

Now we describe the approach with which the classification problem was implemented. Initially, all words in ingredients were lemmatized. After that, embeddings from the fasttext package were taken for all words. Each word was represented by 300-dimensional embeddings. The vector of each ingredient is the sum of the word vectors in that ingredient. For example, the vector for the ingredient "репчатый лук" is the sum of the vectors for the word "репчатый" and the word "лук".

After obtaining the vectors of all the ingredients, we considered two approaches.

Approach 1. Using BoE and a fully connected neural network. Bag-of-Embeddings (BoE) for a recipe is the sum or average of ingredients vectors. The resulting sum (averaging) vector then fed into a fully connected network. We have considered several cases. The first two cases are when the vector is obtained by summing (BoE sum) and taking the average (BoE avg). We also considered cases when weights were added during summation (taking the average), the vectors of ingredients are added with a certain coefficient (BoE sum TF-IDF / BoE avg TF-IDF). These coefficients are TF-IDF.

As a result of the network operation, we have the predicted class. It is the category and cuisine, to which the dish belongs.

Approach 2. Using a convolutional neural network. In a convolutional neural network, we feed the recipe as input as a set of vectors of ingredients. To ensure that all arrays of recipes are the same length, we add empty words until the desired dimension is reached. As a result, we get an array of dimension 302 (the words UNK and PAD are allocated to a different dimension). We traverse these vectors by convolution. For example, in the recipe "яичница" there are ingredients "яйцо", "соль", "масло". Then the visual representation of the convolution with window 2 can be seen in Fig.1.

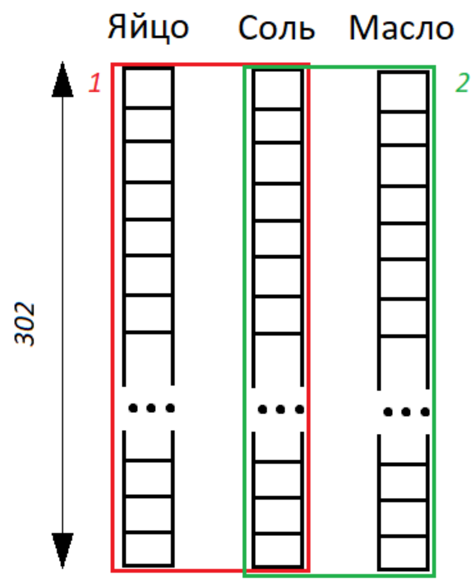

Fig.1. Convolution example with window 2

In this work, three convolutions are used:
1) with window 2;
2) with window 3;
3) with window 4.

The complication of the convolutional approach lies in the fact that the vectors returned by these three convolutions are concatenated into one large vector, which is fed to a fully connected layer (shown schematically in Fig. 2).

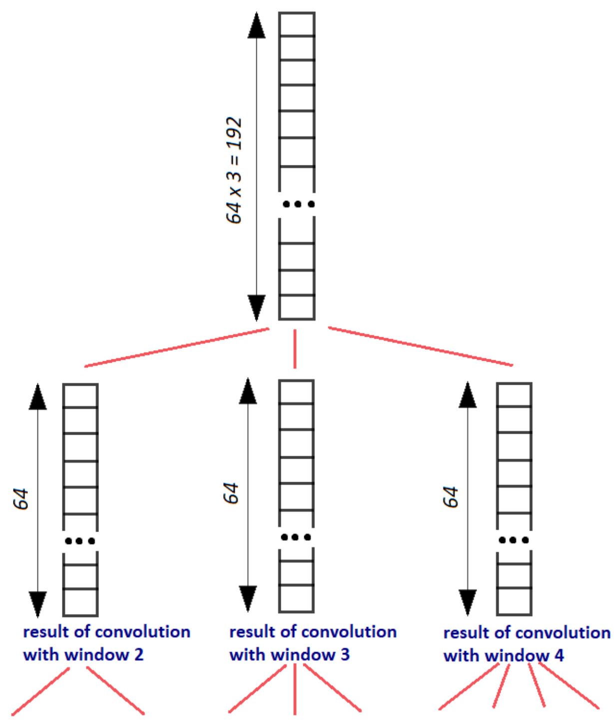

Fig.2. An example of vectors concatenation obtained after convolution

4 Dataset

As mentioned earlier, the existing datasets that are related to the preparation of dishes do not have enough information to solve the problem of belonging to a certain cuisine. We used open Internet sources, which provided descriptions of various dishes. With their help, a Russian-language dataset was assembled which contains the following fields:

- url of the recipe;
- dish name;
- dish category;
- the cuisine this dish belongs to;
- menu type (gluten free, diabetic, unlimited, etc.);
- description of the dish;
- ingredients;
- cooking instructions;
- advices.

Data collection was automated. The program code with which the dataset was uploaded can be viewed a in our repository listed at the abstract. All received data was uploaded to a json file. Such dataset can be well suited for solving various problems in the restaurant business and cooking.

In our task, it was required to use the columns shown in Fig.3.

	id	title	category	cuisine	ingredients
0	28195	Банановое мороженое с корицей	Выпечка и десерты	Карибская кухня	[бананы, лимонный сок, молотая корица, ванильн...
1	28141	Ананасово-кокосовый шербет	Выпечка и десерты	Карибская кухня	[сахар, кокосовое молоко, корень имбиря, анана...
2	27929	Десерт из манго, клубники и текилы	Выпечка и десерты	Карибская кухня	[клубника, манго, сахар, тертая цедра лайма, т...
3	28192	Ананас с соусом из манго и рома	Выпечка и десерты	Карибская кухня	[манго, темный ром, сок лайма, сахар, тертая ц...
4	28463	Клубничный соус с текилой	Выпечка и десерты	Карибская кухня	[клубника, сахарная пудра, сок лайма, текила, ...
...	...	...	...	...	...
41433	43380	Постный борщ с фасолью	Супы	Украинская кухня	[белая фасоль, овощной бульон, свекла, картофе...
41434	80446	Суп энгамат	Супы	Шведская кухня	[цветная капуста, морковь, картофель, лук-поре...
41435	136820	Гороховый суп с блинчиками	Супы	Шведская кухня	[горох, репчатый лук, гвоздика, свиная рулька,...
41436	18014	Суп из брокколи и кростини с сыром бри	Супы	Шведская кухня	[оливковое масло, репчатый лук, овощной бульон...
41437	49530	Сливочный суп с фрикадельками	Супы	Шведская кухня	[картофель, репчатый лук, мясной фарш, морковь...

Fig.3. An example of data from a dataset of recipes

The focus was on the "category", "cuisine" and "ingredients" fields as they are used in our task.

The "category" section contains 9 classes:
1) pastries and desserts;
2) breakfasts;
3) snacks;
4) drinks;
5) main dishes;
6) pasta and pizza;
7) salads;
8) sauces and marinades;
9) soups.

The presented dataset contains 97 different cuisines. As you know, too many classes have a negative impact on the solution of the clustering problem. In this regard, it was decided to reduce the number of classes by combining the cuisines into larger classes. As a result, 13 types of cuisines were obtained, the names of which are presented in the headings of the columns of Tab. 1.

Table 1: Объединение кухонь в более крупные классы.

Балканская	Славянская	Европейская	Юго-Восточная Азия	Восточноазиатская
Албанская	Белорусская	Австрийская	Восточно-индийская	Китайская
Болгарская	Крымская	Бельгийская	Вьетнамская	Корейская
Румынская	Одесская	Британская	Индийская	Японская
Сербская	Русская	Венгерская	Индонезийская	
Турецкая	Советская	Голландская	Камбоджийская	
Хорватская	Украинская	Европейская	Паназиатская	
Югославская		Ирландская	Сингапурская	
Греческая		Испанская	Тайская	
		Итальянская	Филиппинская	

Кавказская	Центральноазиатская	Скандинавская	Африканская
Абхазская	Афганская	Датская	Алжирская
Армянская	Башкирская	Исландская	Африканская
Грузинская	Бурятская	Норвежская	Египетская
Дагестанская	Казахская	Скандинавская	Марокканская
Кавказская	Киргизская	Финская	
Осетинская	Таджикская	Шведская	
Черкесская	Татарская		
Чеченская	Туркменская		
Азербайджанская	Узбекская		

Южноамериканская	Североамериканская	Ближний Восток	Другие кухни
Аргентинская	Американская	Арабская	Мировая
Бразильская	Канадская	Еврейская	Авторская
Карибская	Креольская	Ливанская	Австралийская
Колумбийская	Кубинская	Сирийская	
Латиноамериканская	Мексиканская	Персидская	

5 Experiments

Let us now describe in more detail the features of the software implementation of the algorithm.

5.1 Metrics

Accuracy was chosen as the main metric that we used to evaluate our approaches. This is the ratio of the number of correctly guessed $N_{correctly}$ cuisines/categories to their total number N:

$$Accuracy = \frac{N_{correctly}}{N}.$$

5.2 Experiment Setup

During the implementation of the algorithm using the feedforward network, the ReLu function was used as the activation function, except for the last layer. The activation function on the last layer is softmax.

The categorical crossentropy was used as the Loss function, the optimizer was chosen "rmsprop".

The structure of the neural network and the number of parameters can be seen in the summary, shown in Fig.4.

Layer (type)	Output Shape	Param #
input_4 (InputLayer)	[(None, 302)]	0
dense_6 (Dense)	(None, 64)	19392
dense_7 (Dense)	(None, 16)	1040
dense_8 (Dense)	(None, 96)	1632

Total params: 22,064
Trainable params: 22,064
Non-trainable params: 0

Fig.4. Summary for a model using BoE

Similar parameters were set for the model using a convolutional neural network. The structure of the neural network and the number of parameters can be seen in the summary, shown in Fig.5.

Layer (type)	Output Shape	Param #	Connected to
input_5 (InputLayer)	[(None, 40)]	0	[]
embedding_1 (Embedding)	(None, 40, 302)	1330612	['input_5[0][0]']
conv1d_1 (Conv1D)	(None, 39, 64)	38720	['embedding_1[0][0]']
conv1d_2 (Conv1D)	(None, 38, 64)	58048	['embedding_1[0][0]']
conv1d_3 (Conv1D)	(None, 37, 64)	77376	['embedding_1[0][0]']
global_max_pooling1d_1 (GlobalMaxPooling1D)	(None, 64)	0	['conv1d_1[0][0]']
global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 64)	0	['conv1d_2[0][0]']
global_max_pooling1d_3 (GlobalMaxPooling1D)	(None, 64)	0	['conv1d_3[0][0]']
concatenate (Concatenate)	(None, 192)	0	['global_max_pooling1d_1[0][0]', 'global_max_pooling1d_2[0][0]', 'global_max_pooling1d_3[0][0]']
flatten_1 (Flatten)	(None, 192)	0	['concatenate[0][0]']
dense_9 (Dense)	(None, 32)	6176	['flatten_1[0][0]']
dense_10 (Dense)	(None, 96)	3168	['dense_9[0][0]']

Total params: 1,514,100
Trainable params: 183,488
Non-trainable params: 1,330,612

Fig.5. Summary for the model using CNN

5.3 Baselines

The main idea of the project was borrowed from [1]. Here the authors used fasttext, CNN, Text Recurrent Neural Network (TextRNN), Text bidirectional RNN, Text Attention bidirectional RNN, and Hierarchical Attention Network(HAN). The results of the approaches are demonstrated on an English-language dataset of 39774 records.

We demonstrated similar ideas using BoE and CNN for a self-assembled Russian-language dataset.

6 Results

As a result of the implementation of the approaches described in the previous sections, the following results were obtained (Tab. 2).

Table 2: Results for different models.

	Accuracy	
Model	Category	Cuisine
BoE sum	52.93%	55.38%
BoE avg	26.92%	53.98%
BoE sum TF-IDF	36.06%	53.69%
BoE avg TF-IDF	27.02%	53.18%
CNN	72.70%	58.22%

According to Table 2, the CNN approach showed the best result both for determining the cuisine and for determining the category. For the task of predicting the category, the accuracy reached 72.7%, which can be considered a good result. Among BoE approaches for category and cuisine prediction, the best turned out to be BoE sum.

The results of predicting the category of a dish and the cuisine to which it belongs for the best models are presented below in Fig.6-7.

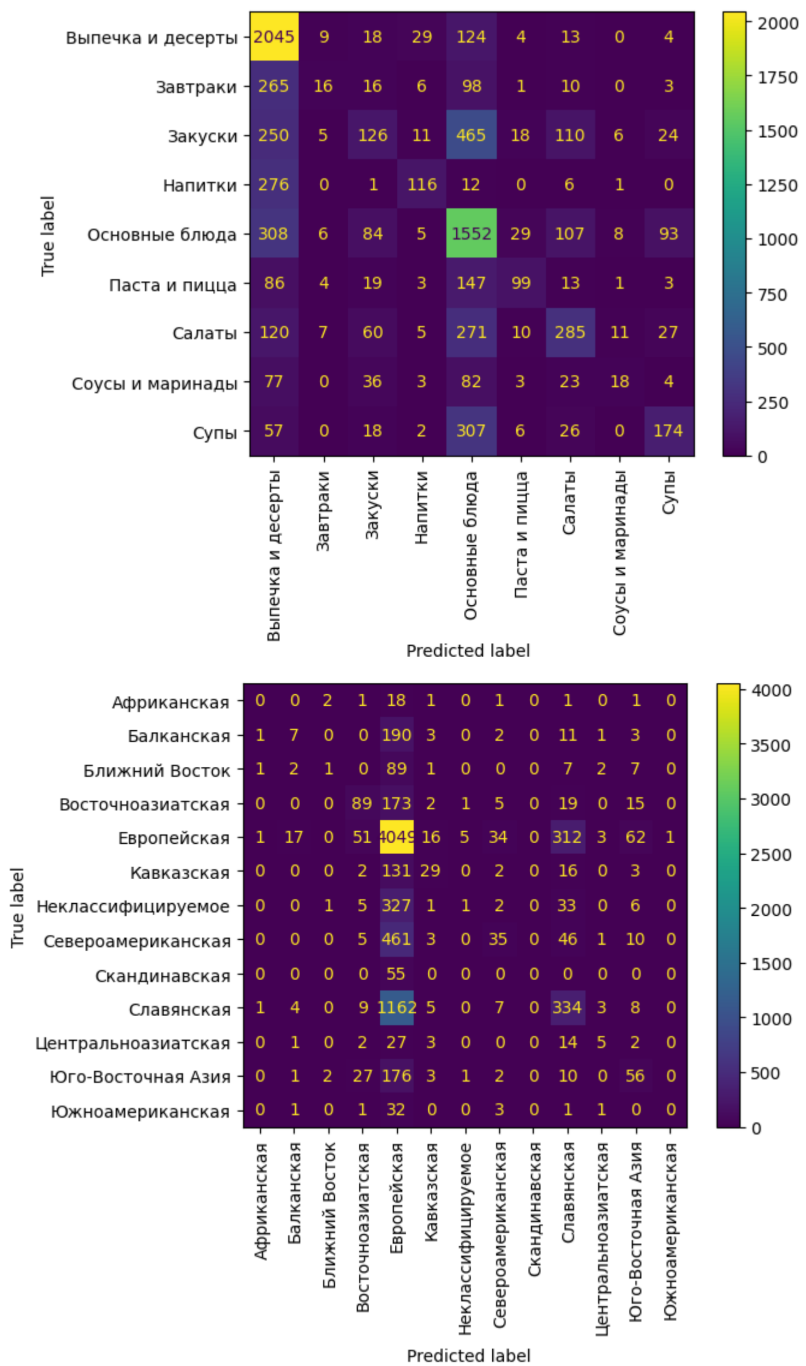

Fig.6. Cuisine and category prediction results for the BoE sum model

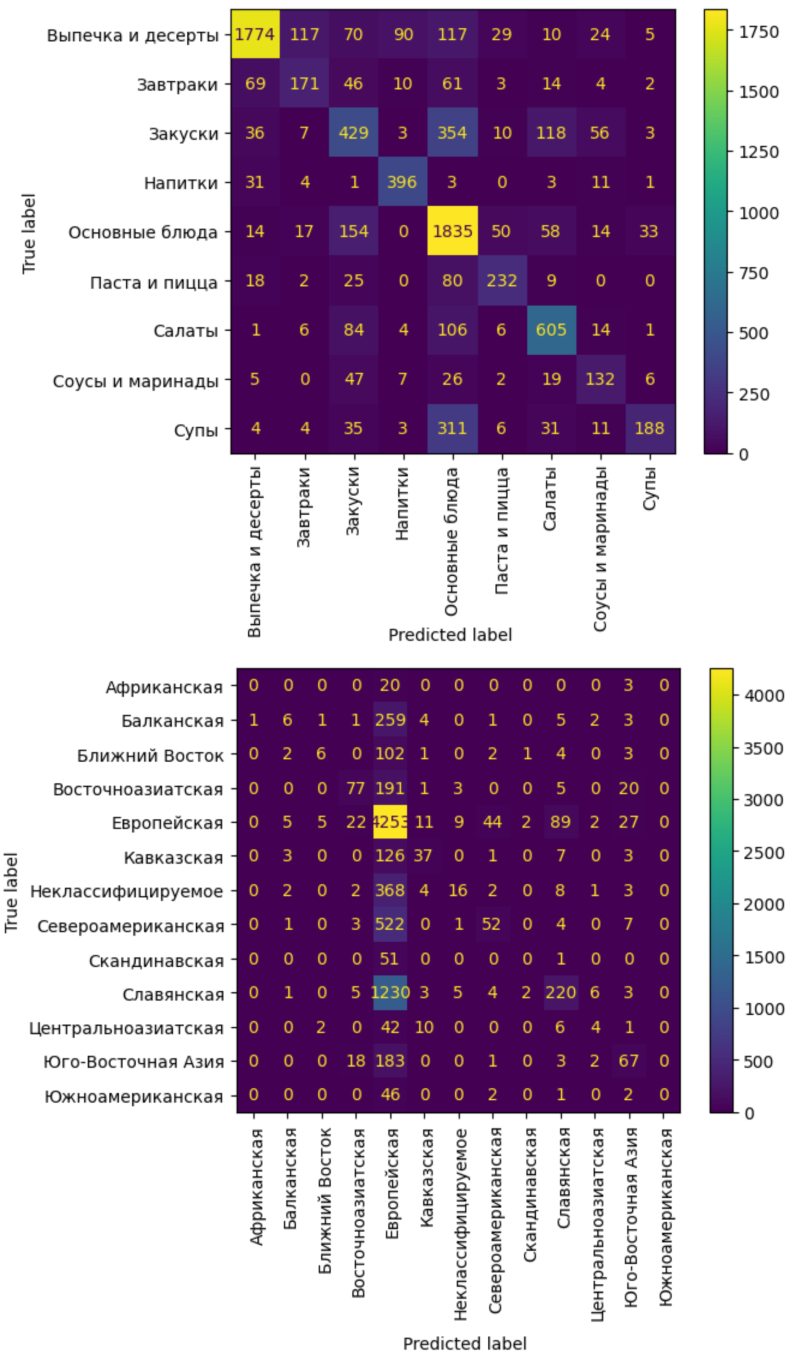

Fig.7. Cuisine prediction results and categories for CNN model

In comparison with the results presented in [1], our cuisine prediction accuracy turned out to be lower. However, this result can be obtained for the

following reasons:
a) the collected dataset is in Russian, approaches should be improved taking this into account;
b) the grouping of 97 kitchens into larger classes was done manually. A different distribution of cuisines into broader classes could change the accuracy of the prediction.

7 Conclusion

In the course of the work done, a Russian-language dataset was collected containing a description of various dishes. You can implement various NLP tasks from this area thanks to this dataset. The main approaches to solving the problem of belonging to a certain category or cuisine were tested on this dataset. In the future these approaches can be modified to improve the prediction accuracy.

References

[1] Ali, A. HAN TO PREDICT CUISINE FROM RECIPES INGREDIENTS / A. Ali, A. Wróblewska // International research journal of science and technology. — 2022. — No. 4(12). — P. 1835-1842.

[2] Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images — URL: http://pic2recipe.csail.mit.edu/ (date of the application: 17.01.2023).

[3] 147к кулинарных рецептов (147k culinary recipes). — // kaggle : [website]. — URL: https://www.kaggle.com/datasets/rogozinushka/povarenok-recipes (date of the application: 26.02.2023).

[4] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.