A Cooking Recipe Multi-Label Classification Approach for Food Restriction Identification

Larissa F. S. Britto[1], **Luciano D. S. Pacifico**[1], **Emilia G. Oliveira**[1], **Teresa B. Ludermir**[2]

[1]Departamento de Computação – Universidade Federal do Rural de Pernambuco
Recife – PE – Brazil

[2]Centro de Informática – Universidade Federal de Pernambuco
Recife – PE – Brazil

{larissa.feliciana,luciano.pacifico,emilia.galdino}@ufrpe.br,tbl@cin.ufpe.br

Abstract. *Recipe sharing websites have become very popular in the past years, allowing individuals to use such systems in an attempt to find a desired recipe. But sometimes finding recipes which best fit the user's wishes, while still satisfying his food restrictions, may become a very time consuming and difficult task. In this work, we propose a recipe multi-label classification approach as part of a recipe recommendation system for people with food restrictions, in an attempt to automatically identify whether an input recipe or list of ingredients fits into one or more food restrictions, satisfying both user's expectations and needs. The experimental evaluation includes two approaches for feature selection, as a manner to reduce the computational costs for the proposed system.*

1. Introduction

Food plays an important hole for human beings, not only as part of its irreplaceable biological functions (such as promoting growth and development, as the main source of energy to the body, and by providing repairing, maintenance and auto-regulation functions), but also as a major anthropological element. One of the most popular ways to learn how to prepare new dishes is the analysis of previously recorded text documents: the cooking recipes.

With the fast advance of information technologies (such as internet, social media and smartphones) in the past few decades, the practice of recipe sharing has been increasing in popularity. Nowadays, enormous recipe repositories can be easily found on the web, and with the relentless globalization process, everyone can have access to a wide variety of ingredients and restaurants from different cultures, countries, ethnicities, culinary styles, and so on [Mokdara et al. 2018]. The choice of the right dish to eat at a certain moment of the day is a hard decision making process, involving many factors like personal tastes, nutritional information concerning the meal, dietary restrictions, allergies, etc. Food choice is also based on the color, texture, sound (i.e., crispness) and the temperature it holds [Nirmal et al. 2018].

But if for one side the popularization of cooking specialized websites has been helping to diffuse the culture of cooking recipe sharing, in the other hand the huge amount of data stored in such repositories makes the searching for a recipe that fits better the user needs and expectations a tough task. In this context of big data systems, Data Mining techniques [Han et al. 2011] (such as Recommendation Systems) have been proposed as computational tools to help users in finding useful and personalized information in their contexts of application.

Recommendation systems [Isinkaye et al. 2015] are automatic filtering techniques employed to reduced the amount of data retrieved by the search of a user. These systems are able to perform personalized information retrieval based on many factors, like the analysis of user profile and history, sentiment and community analysis concerning the product to be recommended, etc. In literature, some cooking recipe recommendation systems have been presented recently [Mokdara et al. 2018, Nirmal et al. 2018, Nezis et al. 2018, Gorbonos et al. 2018, Nilesh et al. 2019, Oliveira et al. 2019, Pacifico et al. 2019, Rong et al. 2019, Zhang et al. 2019]. Recipe recommendation systems could assist users in finding a personalized and balanced diet, promoting good eating habits as a manner to improve their health, once many chronic diseases (such as heart diseases, cancer and diabetes) are associated with bad eating habits [Trattner and Elsweiler 2017].

But sometimes it is hard to find all ingredients from the list of recommended recipes, what makes the preparation of the recommended dish impossible. Also, most part of the recipes contained in such repositories do not contemplate people who have some kind of food restriction, like food-allergenic patients and some cultural groups (i.e, vegans, vegetarians, some religious groups, and so on) [Ooi et al. 2015].

In this work, as part of an effort towards the development of more precise recipe recommendation systems that would contemplate users with some kind of food restriction, a multi-label text classification approach is proposed to automatically identify recipes that would fit one or more categories of food restrictions, to compose the recognition/decision module of such systems. The proposed method is a data-driven approach, where recipe documents are obtained from specialized websites and used to train machine learning algorithms. In an attempt to reduce the computational costs for the proposed recommendation systems, two feature selection approaches are implemented, seeking out to find the best set of ingredients that would be necessary to detect a specific recipe category.

The work is organized as follows. Firstly, a brief discussion on some related works on cooking recipe classification are presented (Section 2). The proposed methodology is discussed in Section 3, where the data set acquisition process (Section 3.1) and the selected classifiers (Section 3.2) are briefly described. After that, our three-steps experimental methodology is presented, as much as the analysis on the experimental results (Section 4). Finally, some conclusions and leads for future works are shown (Section 5).

2. Related Works

Some cooking recipe classification works from the literature are briefly described as follows.

Su et al. [Su et al. 2014] evaluated the correlations between recipe cuisines and their ingredients in an attempt to investigate the underlying cuisine-ingredient connections. Associative classification and Support Vector Machine have been used as the recognition modules, and the authors attempted to explain the correlation among different recipe cuisines by means of the identification of the ingredients belonging in each cuisine and the confusion matrix generated by the classifiers.

Jayaraman et al. [Jayaraman et al. 2017] also attempted to offer a comprehensive analysis underlying the correlation between recipe cuisines and their list of ingredients. The proposed methodology evaluated the performances of four classification algorithms (Naive Bayes, Multinominal Logistic Regression, Random Forest Classifier and Linear Support Vector Machine) when dealing with recipe classification problem. The authors applied some text pre-processing techniques to standardize the data set adopted for experimental analysis.

In Nirmal et al. [Nirmal et al. 2018], a recipe recommendation system is proposed based on the optimization of both flavor and nutritional value of the ingredients. A Random Forest Classifier algorithm is employed to perform the automatic classification of recipe in cuisines as the first step of the proposed system.

In Kalajdziski et al. [Kalajdziski et al. 2018], some well-established text pre-processing and feature selection methods (like TF-IDF and Bag of Words techniques) have been employed to perform automatic recipe cuisine classification. The Naive Bayes classifier, an Artificial Neural Network and the Support Vector Machine algorithms were adopted as the recognition module, and the best combination of feature selection and classifier has been selected after the experiments as the resulting system.

In [Britto et al. 2019] et al., a complete recipe classification system is proposed and evaluated for Brazilian Portuguese recipe documents. Several classifiers from Machine Learning literature (including a Multi-Layer Perceptron) were tested as the recognition module for the proposed system, and a new data set (composed of more than 3000 recipes, and more than 1300 ingredients) was presented.

3. Methodology

The methodology is composed of two parts: data set acquisition (Section 3.1) and the definition of the classifier (Section 3.2). Each part is described as follows.

3.1. Data Set

In this work, the recipe data set proposed in [Majumder et al. 2019] is selected for experimental purposes. This data set has been extracted from Food.com[1] website, and made available by the authors[2]. The original data set contains more than 230,000 recipes in English, and more than 1,000,000 user interactions with the website, between the years 2000 and 2018.

In our approach, only the list of ingredients and the categories (extracted from the list of user tags) for each recipe are taken into consideration, in a methodology similar to many works in recipe classification literature [Jayaraman et al. 2017, Su et al. 2014, Kalajdziski et al. 2018, Nirmal et al. 2018]. In this work, we are interested in recipes that are related to some kind of food restriction only, so ten recipe categories are proposed, and the original user tags are distributed into the proposed categories (Table 1). Once many recipes from the original data set are not labeled, and many of the original tags are not related to food restrictions, a total of 29951 documents is obtained, and the final list of ingredients is composed of 6734 different ingredients.

[1] `https://www.food.com/`

[2] `https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions`

Tabela 1. Proposed class labels and the corresponding original set of user tags.

Class Name	Original Tags
pork	bacon, ham, pork
nuts-grains	bean, nuts, grains
meat	beef, chicken, meat, steak, turkey, lamb, quail, rabbit, poultry
gluten	bread, pasta, spaghetti, macaroni, penne
fish	fish, cod, salmon, halibut, trout, seafood
dairy	cheese, lactose
seafood	crab, lobster, octopus, oysters, scallops, squid
restrict diet	dairy-free, gluten-free, hanukkah, high-calcium, high-fiber, high-protein kosher, jewish, low-carb, low-fat, low-protein, low-sodium, hashana vegan, vegetarian, healthy, diabetic
egg	egg, omelet
shrimp	shrimp

Tabela 2. Data distribution among the recipe categories: the main diagonal keeps the total number of recipes per category.

Class	1	2	3	4	5	6	7	8	9	10
1	**1415**	197	1415	151	14	177	16	592	236	6
2	197	**6610**	663	4238	227	702	234	4019	1214	119
3	1415	663	**4839**	549	33	401	41	2070	504	15
4	151	4238	549	**7080**	201	765	205	4263	1238	106
5	14	227	33	201	**1384**	84	1384	649	129	465
6	177	702	401	765	84	**2843**	87	1789	2510	21
7	16	234	41	205	1384	87	**1445**	673	132	465
8	592	4019	2070	4263	649	1789	673	**22243**	3561	211
9	236	1214	504	1238	129	2510	132	3561	**5482**	38
10	6	119	15	106	465	21	465	211	38	**465**

The cooking recipe classification problem can be easily mapped into a standard text classification problem [Jayaraman et al. 2017, Kalajdziski et al. 2018], using the following assumptions:

- Data Pattern = Document = Recipe;
- Feature = Word = Ingredient;
- Class = Document Class = Recipe Category.

This work employs a multi-label classification approach for the categorization of the recipe documents, that is, the recipes may be attributed to more than one category at the same time. Due to the fact that many documents are labeled by the users using more than one tag, the standard multi-class categorization methodology seems to be too much restrictive, not representing the best way to deal with the classification problem, in this case. Table 2 presents the data distribution among the proposed recipe categories and the overlapping among such categories. As we can observe, some of the adopted categories are sub-categories from others, but we opted to keep than separated according to the actual problem restrictions. For instance, *fish* category is a sub-category of *seafood* according to the original user tags, but users that have no restriction to eat fish may have some restrictions in relation to other kinds of seafood.

Considering the previous assumptions, the data set of recipe documents is converted into a Document-Term Matrix (DTM) [Jayaraman et al. 2017, Kalajdziski et al. 2018], so we could use than as the input for the recognition module of the proposed system. The obtained DTM is a (29951×6734)-dimensional binary-coded sparse matrix, where each row corresponds to a recipe and each column corresponds to an ingredient. In a similar fashion, the Target Matrix (TM) is a (29951×10)-dimensional binary-coded matrix.

3.2. Random Forest Classifier

In this work, we adopt the Random Forest Classifier (RFC) algorithm [Criminisi et al. 2011] as the recognition module for our proposed multi-label recipe classification system. Random Forest Classifiers is a robust supervised method from Machine Learning literature, and it has been successfully applied to deal with recipe classification task recently [Nirmal et al. 2018]. In RFC, a set of Decision Tree Classifiers (DTCs) [Mitchell et al. 1997] is implemented as the core decision-making mechanism. The Random Forest Classifier is built as an ensemble learning method which combines the prediction of many DTCs (using some mechanism, such as majority voting). RFC creates a set of decision trees from randomly selected sub-sets of the training data, aggregating the votes from different estimators (i.e., DTCs) to decide the final class of a given testing sample.

For the proposed multi-label classification approach, a different RFC will be trained to identify each one of the data classes, in a One-vs-All (also known as One-vs-Rest) methodology [Bishop 2006].

4. Experimental Evaluation

In this section, we test the proposed multi-label approach for food restriction identification in recipes by means of three sets of experiments. Firstly, the whole data set is tested, so we could find the best number of estimators for the Random Forest Classifier. The second evaluation step is performed by using the Bag of Words approach to select the best set of most frequent features considering the whole data set, in an attempt to reduce the number of required features to train the RFCs. Finally, in the third part of the evaluation, Bag of Words is also used to select the best number of most frequent features, but this time, the selection is performed for each recipe category, once more in an attempt to reduce the final number of features required to train the RFC.

All experiments have been implemented in Python programming language, and all tests have been executed in a computer with an i7-7700K CPU, NVIDIA GeForce GTX 1060 6GB GPU and 32 GB RAM. The RFC has been implemented using scikit-learn library [Pedregosa et al. 2011, Buitinck et al. 2013]. For evaluation purposes, three well-established classification metrics are employed: F-Measure (eq. (1)), Precision (eq. (2)) and Recall (eq. (3)).

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{1}$$

where

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{2}$$

and

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

Our experiments have been conducted using a k-Folds Cross-Validation framework, with ten folds. The data set has been randomly split into ten parts to form the training and testing sets. Nine folds are used each time to compose the training set, and the remaining fold is used as the testing set. For all three experimental scenarios, the

Tabela 3. Experimental results for whole data set and variant number of estimators for the Random Forest Classifier: Mean $\pm$ standard deviation. The training time is reported in seconds.

N. of Est.	F-Measure	Precision	Recall	Training Time
10	0.6957 $\pm$ 0.0074	0.7711 $\pm$ 0.0076	0.6918 $\pm$ 0.0086	172.4462 $\pm$ 2.4900
20	0.7144 $\pm$ 0.0065	0.7843 $\pm$ 0.0070	0.7149 $\pm$ 0.0068	346.2110 $\pm$ 5.0065
30	0.7197 $\pm$ 0.0056	0.7866 $\pm$ 0.0075	0.7227 $\pm$ 0.0055	519.4476 $\pm$ 5.9923
40	0.7236 $\pm$ 0.0080	0.7908 $\pm$ 0.0081	0.7264 $\pm$ 0.0087	693.5489 $\pm$ 6.7611
50	0.7260 $\pm$ 0.0061	0.7918 $\pm$ 0.0062	0.7304 $\pm$ 0.0068	857.6620 $\pm$ 5.4625
60	0.7259 $\pm$ 0.0071	0.7912 $\pm$ 0.0087	0.7300 $\pm$ 0.0072	1028.1 $\pm$ 5.8250
70	0.7276 $\pm$ 0.0073	0.7930 $\pm$ 0.0082	0.7319 $\pm$ 0.0083	1199.9 $\pm$ 7.9784
80	**0.7286** $\pm$ **0.0075**	0.7937 $\pm$ 0.0088	0.7336 $\pm$ 0.0076	1375.4 $\pm$ 10.6508
90	**0.7286** $\pm$ **0.0080**	0.7934 $\pm$ 0.0094	**0.7337** $\pm$ **0.0074**	1552.4 $\pm$ 14.6247
100	**0.7286** $\pm$ **0.0063**	**0.7938** $\pm$ **0.0076**	**0.7337** $\pm$ **0.0063**	1703.6 $\pm$ 11.2494

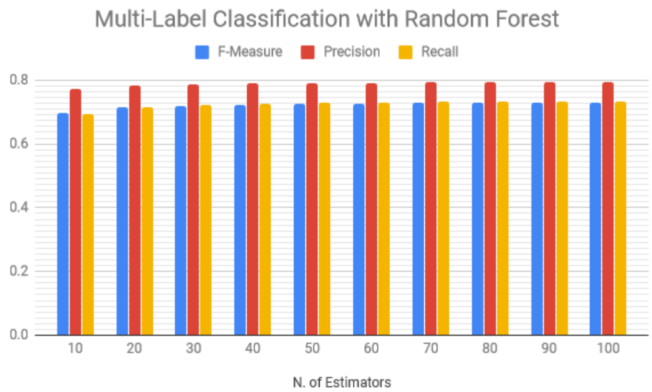

Figura 1. Average values according to the selected number of estimators for Random Forest Classifier, considering the whole data set.

data patterns in each fold have been kept the same, so we could perform a fair evaluation on the actual influence of the parameter in analysis (number of estimators for the RFC, or the set of selected features) each time.

The experimental results for the first step of our evaluation are presented in Table 3. In this step, the whole set of features (i.e., 6734 ingredients) is evaluated, in an effort to find the best number of estimators for the RFC. As we can observe (Fig. 1), the best values for all three metrics have been found by the RFC using 100 estimators, which also presented the best degree of stability. Therefore, the second and third steps for current evaluation will be carried out using a RFC with 100 estimators. Table 3 also showed that the selected indices did not improve very much significantly for a number of estimators greater than 80, so if time is a critical factor, one should take this trade-off into consideration, once the training time increases considerably when the number of RFC estimators is high.

Seeking out to reduce the computational costs for training the RFCs and the need for storage memory, the second step for our experimentation is driven in an attempt to reduce the feature set (i.e., the total number of ingredients) required to perform a good classification of the data patterns (recipes) into the selected set of classes (recipe categories). At first, we will use Bag of Words to select the top most frequent ingredients in relation to the whole data set (Fig. 2), so we could identify the best reduced set of ingredients that are necessary to train the RFCs properly for each recipe category.

Figura 2. Most frequent ingredients considering the whole data set.

Tabela 4. Experimental results for Bag of Words considering the most frequent ingredients for the whole data set and a RFC using 100 estimators: Mean $\pm$ standard deviation. The training time is reported in seconds.

N. Feat.	F-Measure	Precision	Recall	Training Time
300	0.6540 ± 0.0085	0.7264 ± 0.0096	0.6608 ± 0.0085	92.1084 ± 0.9422
600	0.6861 ± 0.0085	0.7499 ± 0.0083	0.6967 ± 0.0101	202.6036 ± 0.5066
900	0.7024 ± 0.0073	0.7627 ± 0.0080	0.7147 ± 0.0091	337.5291 ± 1.4796
1200	0.7121 ± 0.0061	0.7704 ± 0.0058	0.7252 ± 0.0082	460.8004 ± 2.1311
1500	0.7187 ± 0.0066	0.7756 ± 0.0074	0.7330 ± 0.0070	593.1686 ± 4.1092
1800	0.7235 ± 0.0064	0.7796 ± 0.0070	0.7380 ± 0.0077	714.9733 ± 3.6605
2100	0.7268 ± 0.0074	0.7824 ± 0.0082	0.7415 ± 0.0085	822.9208 ± 6.9074
2400	0.7271 ± 0.0078	0.7840 ± 0.0086	0.7399 ± 0.0087	919.9190 ± 8.5347
2700	$\mathbf{0.7272 \pm 0.0059}$	$\mathbf{0.7847 \pm 0.0071}$	$\mathbf{0.7391 \pm 0.0071}$	1011.1 ± 6.4307
3000	0.7261 ± 0.0073	0.7844 ± 0.0079	0.7372 ± 0.0082	1092.3 ± 10.0388

The experimental results for the global feature selection are presented in Table 4. In this step, the RFC with 100 estimators has been used for evaluation, according to the experimental analysis performed on the first step of current investigation. As pointed out by the results (Fig. 3), the best set of features is composed of 2700 ingredients, and there is no significant improvement in relation to the selected classification indices for values greater than 2700 features. The original data set is composed of documents submitted by the users of the website, which in general, are too much specific when including ingredient names to their recipes, making the list of ingredients with low occurrences too large (about 2300 ingredients occur in only one recipe of the selected data set), what makes the training process too slow. As can be observed, the final list contains less than half of the original ingredients, but the obtained values for the classification indices are very close to the values obtained by the tests using the whole data and ingredient sets. But the main contributions of the feature selection process can be noticed when we compare the training times and the dimensions of the resulting DTMs: the original DTM is a (29951×6734)-dimensional matrix, and the RFC has spent an average time of about 1703 seconds to training in each experiment execution, while the reduced DTM is a (29951×2700)-dimensional matrix, leading to an average training time of about 1011 seconds.

The third step of the experimentation has been performed by the application of Bag of Words to select the best set of ingredients for each one of the proposed recipe categories (Fig. 4). This evaluation is performed in an attempt to avoid the influence of

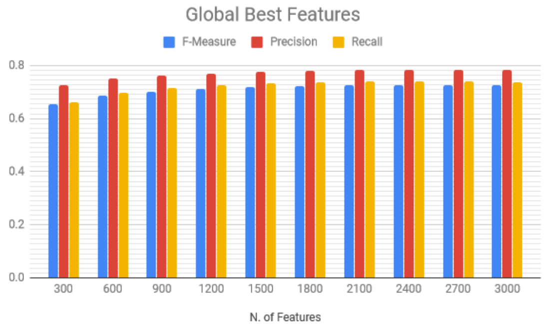

Figura 3. Average values according to the total number of most frequent ingredients.

Tabela 5. Experimental results for Bag of Words considering the most frequent ingredients per recipe category and a RFC using 100 estimators: Mean ± standard deviation. The training time is reported in seconds.

Feat. per Category	Total N. Feat.	F-Measure	Precision	Recall	Training Time
100	256	0.6739 ± 0.0077	0.7425 ± 0.0080	0.6821 ± 0.0089	92.0468 ± 0.6675
200	481	0.6989 ± 0.0064	0.7609 ± 0.0056	0.7094 ± 0.0087	173.5610 ± 1.0019
300	726	0.7130 ± 0.0068	0.7715 ± 0.0078	0.7259 ± 0.0077	281.4169 ± 1.8973
400	980	0.7206 ± 0.0074	0.7764 ± 0.0066	0.7353 ± 0.0091	409.9541 ± 2.6786
500	1244	0.7239 ± 0.0072	0.7793 ± 0.0079	0.7386 ± 0.0085	530.3208 ± 2.7339
600	1524	0.7282 ± 0.0069	0.7831 ± 0.0088	0.7430 ± 0.0070	647.9850 ± 4.7345
700	1753	0.7282 ± 0.0062	0.7838 ± 0.0077	0.7425 ± 0.0065	720.9341 ± 4.9627
800	1973	0.7294 ± 0.0062	0.7849 ± 0.0065	0.7435 ± 0.0073	812.1686 ± 4.8073
900	2240	**0.7306 ± 0.0072**	**0.7866 ± 0.0075**	**0.7440 ± 0.0078**	884.2394 ± 10.1850
1000	2540	0.7297 ± 0.0070	**0.7866 ± 0.0084**	0.7427 ± 0.0068	975.4369 ± 9.1505

class dominance among each other, due to the fact that the selected data set is unbalanced in relation to the recipe categories.

The experimental results for the third and last set of experiments are shown in Table 5. Once more, a RFC with 100 estimators has been trained as the classifier for each recipe category. As showed by the experimental results (Fig. 5), this experimental testing bed has presented the best average performances concerning the selected metrics in comparison to the other two testing scenarios. When considering the most frequent ingredients in each recipe category, some ingredients that would hardly reach high ranks from a global point of view, but that are quite relevant for specific recipe categories, have the opportunity to be included in the final feature set (see Fig. 4), leading to better classification performances.

The best results have been found for 900 ingredients per category. As can be observed (Fig. 6), the expected number of features is 9000 (although the whole data set is composed of only 6734 unique ingredients), but the actual number of selected features is 2240, what indicates that there is a huge overlapping related to the most frequent ingredients through multiple recipe categories. Once more, the resulting DTM dimensions and average training time are very advantageous in comparison to the use of the complete set of unique ingredients (feature set).

In an overall evaluation, we conclude that the best evaluation scenario has been obtained when the 900 most frequent ingredients from each category are used, resulting

Figura 4. Most frequent ingredients for each recipe category: (a) Pork, (b) Nut and Grains, (c) Meat, (d) Gluten, (e) Fish, (f) Dairy, (g) Restricted Diet, (h) Egg, (i) Shrimp.

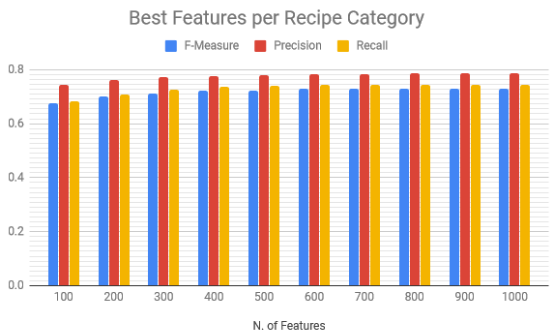

Figura 5. Average values according to the number of most frequent ingredients per recipe category.

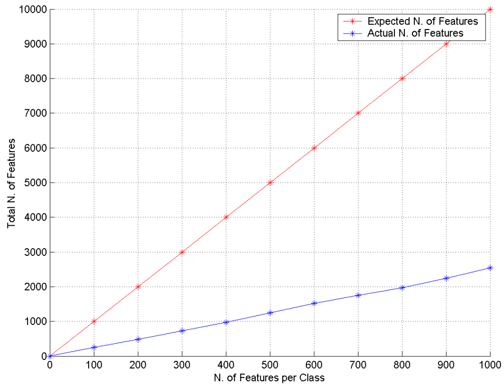

Figura 6. Expected number of features (number of features per class x number of classes) x Actual number of features: for each number of selected features per category, about 75% of the ingredients overlap among the classes.

in a (29951×2240)-dimensional DTM and an average training time of about 884 seconds in each experiment for the RFC with 100 estimators.

5. Conclusions

In this paper, a multi-label approach for cooking recipe classification is presented, for the automatic categorization of food restriction in cooking recipes. The automatic classification of recipes is a fundamental step towards the development of more precise recipe recommendation systems, which may be used for many purposes, such as diet planning, nutritional analysis, recipe generation, and so on.

The proposed approach is implemented as a three-steps evaluation methodology. Firstly, we tested Random Forest Classifier (RFC) in an attempt to select the best number of estimators for the model. The second step consisted in an attempt to select the best set of features (recipe ingredients) globally, taking into consideration the frequency of each ingredient, according to the Bag of Words method. Finally, the third step of the evaluation consisted on the selection of the most frequent ingredients in a local perspective, that is, considering the most frequent ingredients in each recipe category. All evaluations have been performed according to a k-Folds Cross-Validation framework, using three classification indices (F-Measure, Precision and Recall) and the average training time as the validation metrics.

The experimental results showed that the most successful approaches have considered the best set of ingredients for the categories (the local selection methodology). It is completely understandable, once some sets of ingredients may be used to identify a recipe category with high probabilities. The resulting set of selected features found by the experimentation represents an advantage to the task of multi-label classification, once the evaluated data set has been reduced by less than half, keeping the good classification performances, and likewise, reducing the amount of time required for training the adopted classification algorithm.

As future works, we intend to perform a deeper evaluation on the feature set in an effort towards the identification of the best set of ingredients that would be sufficient to detect a recipe category, reducing the computational costs for the recommendation system. This evaluation will be conducted automatically, using techniques such as the Evolutionary Algorithms and other meta-heuristics. We also intend to develop a complete recommendation system for users that present some kind of food restriction, in a way that such users would benefit from the use of these systems, by receiving recommendations for complete and well-balanced diets that are adequate according to their restrictions and tastes.

Acknowledgment

The authors would like to thank FACEPE, CNPq and CAPES (Brazilian Research Agencies) for their financial support.

Referências

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Britto, L. F. S., Oliveira, E. G., Pacifico, L. D. S., and Ludermir, T. B. (2019). A text analysis approach for cooking recipe classification based on brazilian portuguese documents. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)*, 1:436–447.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122.

Criminisi, A., Shotton, J., and Konukoglu, E. (2011). Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning [internet]. *Microsoft Research*.

Gorbonos, E., Liu, Y., and Hoàng, C. T. (2018). Nutrec: Nutrition oriented online recipe recommender. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 25–32. IEEE.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261–273.

Jayaraman, S., Choudhury, T., and Kumar, P. (2017). Analysis of classification models based on cuisine prediction using machine learning. In *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, pages 1485–1490. IEEE.

Kalajdziski, S., Radevski, G., Ivanoska, I., Trivodaliev, K., and Stojkoska, B. R. (2018). Cuisine classification using recipe's ingredients. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1074–1079. IEEE.

Majumder, B. P., Li, S., Ni, J., and McAuley, J. (2019). Generating personalized recipes from historical user preferences. *arXiv preprint arXiv:1909.00105*.

Mitchell, T. M. et al. (1997). Machine learning. wcb.

Mokdara, T., Pusawiro, P., and Harnsomburana, J. (2018). Personalized food recommendation using deep neural network. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*, pages 1–4. IEEE.

Nezis, A., Papageorgiou, H., Georgiadis, P., Jiskra, P., Pappas, D., and Pontiki, M. (2018). Towards a fully personalized food recommendation tool. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, page 77. ACM.

Nilesh, N., Kumari, M., Hazarika, P., and Raman, V. (2019). Recommendation of indian cuisine recipes based on ingredients. In *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, pages 96–99. IEEE.

Nirmal, I., Caldera, A., and Bandara, R. D. (2018). Optimization framework for flavour and nutrition balanced recipe: A data driven approach. In *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pages 1–9. IEEE.

Oliveira, E. G., Britto, L. F. S., Pacifico, L. D. S., and Ludermir, T. B. (2019). Recipe recommendation and generation based on ingredient substitution. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)*, 1:238–249.

Ooi, A., Iiba, T., and Takano, K. (2015). Ingredient substitute recommendation for allergy-safe cooking based on food context. In *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pages 444–449. IEEE.

Pacifico, L. D. S., Oliveira, E. G., Britto, L. F. S., and Ludermir, T. B. (2019). Sistemas de recomendação e geração de receitas através da categorização ontológica dos ingredientes. In *Symposium in Information and Human Language Technology (STIL 2019)*, volume 1, pages 81–85. SBC.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Rong, C., Liu, Z., Huo, N., and Sun, H. (2019). Exploring chinese dietary habits using recipes extracted from websites. *IEEE Access*, 7:24354–24361.

Su, H., Lin, T.-W., Li, C.-T., Shan, M.-K., and Chang, J. (2014). Automatic recipe cuisine classification by ingredients. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*, pages 565–570. ACM.

Trattner, C. and Elsweiler, D. (2017). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498. International World Wide Web Conferences Steering Committee.

Zhang, L., Zhao, J., Li, S., Shi, B., and Duan, L.-Y. (2019). From market to dish: Multi-ingredient image recognition for personalized recipe recommendation. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1252–1257. IEEE.