Supplementary - INDoRI: Indian Dataset of Recipes and Ingredients and its Ingredient Network

Sandeep Khanna[1], Chiranjoy Chattopadhyay[2], Suman Kundu[1]

[1] Indian Institute of Technology Jodhpur,
{khanna.1,suman}@iitj.ac.in
[2] FLAME University, Pune
chiranjoy.chattopadhyay@flame.edu.in

1 Quality Assessment of ISW

Stop words for food are being proposed in order to create an ingredient network. So that clean ingredient names can be scraped from the ingredient list, and thus for universal acceptance, it is necessary to test its usefulness on other cuisines as well. We have tested for other global cuisine including Italian [1], Japanese [2], American [3]. For each of the cuisine hundred recipes were taken along with the ingredients needed to prepare them. The ingredient names were extracted manually and through stop word removal using ISW.

Let $R100 = \{R_1, R_2, \ldots, R_n\}$ represents the set of hundred recipes of each cuisine, $CIM = \{CI_1, CI_2, \ldots, CI_t\}$ represents the set of clean ingredients for some recipe in the set $R100$ that were extracted manually. Here t denotes maximum ingredient count for any recipe, $CINW = \{CI_1, CI_2, \ldots, CI_s\}$ represents the set of clean ingredients for a recipe in the set $R100$ that were extracted using novel words.

The accuracy (similarity between manually extracted ingredients and ingredients extracted using novel stop words) for each recipe of a particular cuisine is calculated using:

$$AccR_i = \frac{|CIM \cap CINW|}{|CIM|} * 100 \tag{1}$$

The average accuracy of a particular cuisine C_i is then calculated using the below formula.

$$AvgAccC_i = \frac{1}{100} \sum_{n=1}^{100} AccR_n \tag{2}$$

For Indian cuisine the average accuracy is over 80 percentage. The results were also promising for Italian cuisines ranging over 70 percentage whereas for Japanese recipes the results were above average with accuracy ranging between 50 and 55.

2 Results and Analysis

We have shown the results on two described applications: category classification and cuisine classification on top of INDoRI. The outcomes of these tasks are elaborated in the following Sections.

2.1 Recipe Categorization

INDoRI contains 8 categories of recipes, encompassing Breakfast, Bread, Dessert, Drink, Chutney, Raita, Lunch/Dinner, and Pickle. We have extracted embeddings of cooking instructions and the ingredients individually. These embeddings were combined to create the final embeddings.

We employed both conventional machine learning techniques and state-of-the-art deep learning models, including advanced transformer-based models. The results are detailed in Table 1 Task 1, presenting the obtained outcomes. It is noteworthy that the traditional machine learning models outperform their deep learning counterparts in this context. Notably, Gradient Boosting achieves the highest F1 score of 0.76, closely followed by Random Forest. Among the sequence-to-sequence models, Bidirectional-LSTM demonstrates a commendable F1 score of 0.66.

2.2 Cuisine Classification

INDoRI encompasses 18 distinct cuisines or regions. The previously computed final embeddings were employed to ascertain the cuisine types. Similarly, the same set of models was utilized for this specific task. The results are depicted in Table 1 Task 2, illustrating the attained outcomes. Once again, the traditional machine learning models outshine their deep learning counterparts in this context. Notably, both Gradient Boosting and Random Forest achieve the highest F1 score of 0.41. In contrast, the deep learning models report F1 scores below 0.50. Notice that the performance for the cuisine classification is substantially lower as some cuisines have comparatively few samples. This points to interesting challenges for few shot learning in NLP.

Table 1: Category vs Cuisine Classification: Task 1 represents classification metrics on Category and Task 2 represent classification metrics on Cuisine.

Task	**Task 1**			**Task 2**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Machine Learning Models						
SVM (RBF Kernel)	0.72	0.57	0.62	0.46	0.25	0.32
Random Forest	0.76	0.78	0.75	0.48	0.43	0.41
GradientBoosting	0.77	0.77	0.76	0.43	0.43	0.41
Sequence to Sequence Models						
LSTM [4]	0.53	0.43	0.46	0.15	0.32	0.27
Bidirectional-LSTM [5]	0.70	0.64	0.66	0.41	0.34	0.37
Transformer Models						
BERT [6]	0.49	0.55	0.43	0.18	0.28	0.21
RoBERTa [7]	0.27	0.52	0.35	0.17	0.25	0.16

References

1. tasteatlas, "100 most popular italian dishes," https://www.tasteatlas.com/100-most-popular-dishes-in-italy [accessed June 6, 2021].
2. deliciousteam, "100 japanese recipes that aren't all sushi," https://www.delicious.com.au/recipes/collections/gallery/55-quick-and-easy-japanese-recipes-to-try-tonight/l1y6brzi [last visited June 2021]., 2021.
3. deliciousteam, "100 most popular american recipes," https://www.deliciousmagazine.co.uk/cuisine/american-recipes [accessed June 5, 2021].
4. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
5. A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional lstm networks for improved phoneme classification and recognition," in *International conference on artificial neural networks*, pp. 799–804, Springer, 2005.
6. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
7. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.