

Kissing Cuisines:

Exploring Worldwide Culinary Habits on the Web

Sina Sajadmanesh Hamid R. Rabiee Mirco Musolesi Sina Jafarzadeh Hamed Haddadi Emiliano De Cristofaro Seyed Ali Ossia Yelena Mejova Gianluca Stringhini

Yummly



Dataset

Dataset

Yummly

- Over 157K recipes in more that 200 cuisines (extracted from Wikipedia).
- More than 3K unique ingredients after standardization process.
- Recipes contain ingredients, flavors, nutrition data, and user ratings.
- BBC Food
 - More than 1000 unique ingredients.
 - Used for ingredient standardization process.
- **D** Country Health Statistics
 - Diabetes prevalence estimates from the World Bank.
 - Health expenditure as a fraction of GDP from the World Bank.
 - Obesity prevalence from World Health Organization.

Characterizing Cuisines

Diversity of Ingredients

Global Diversity

- What is the number of ingredients present in a certain cuisine?
- Local Diversity
 - Do different dishes usually share ingredients or have different sets of ingredients?





Figure 1: Diversity of ingredients around the world

Complexity of Dishes



0.0146 0.0148 0.0150 0.0152 0.0154 0.0156 0.0158 0.0160

Figure 4: Complexity of dishes around the world

Notable Ingredients

- Ingredients as cuisine signatures
 - Some ingredients are widely used, some others are local to specific cuisines.
 - Notable ingredients tend to signify different cuisines.
 - A TF-IDF based method is used to recognize notable ingredients.



Figure 5: Notable ingredients in Italian, Indian, and Mexican cuisines. More notable ingredients have been drawn larger.

Similarity between Cuisines

Similarity of Cuisines

- Similarity point of views:
 - Ingredient-based similarity
 - Flavor-based similarity
- Ingredient-based similarity:
 - Each cuisine is modeled as a distribution over the ingredients.
 - Jensen-Shannon divergence is used to measure the similarity: $JS(P,Q) = \frac{1}{2} [KL(P \parallel M) + KL(Q \parallel M)], \qquad M = \frac{1}{2} (P + Q)$
- **•** Flavor-based similarity:
 - Each cuisine is modeled as a multivariate Gaussian distribution over the flavors.
 - Symmetrized Kullback-Leibler divergence is used as a distance metric:

 ¹/₂ [KL(P || Q) + KL(Q || P)]

Similarity of Cuisines



Figure 6: Graph of similarity between different cuisines in terms of their ingredients and flavors. Each cuisine is linked with five most similar ones. Color of a cuisines denote the geographical region it resides in.

Cuisine Classification

Cuisine Classification

Task

- Assigning a recipe to a cuisine (or a region) based on its ingredients.
- Classifiers
 - Support Vector Machines
 - Deep Neural Networks



Figure 7: The prediction performance of different methods for cuisine and region prediction tasks.

Health and Nutrition

Health and Nutrition

Nutrients:

- Carbohydrate
- Calorie
- Fat
- Protein
- Sugar

Health Statistics:

- Diabetes
- Obesity
- Health Expenditure

Table 2: Correlation of Different Health Measures with Nutrition Values of Recipes

		Correlation Values	
Health Measure	Nutrient	Pearson	Kendall-Tao
Obesity	Calorie	-0.104	-0.110
	Protein	-0.483	-0.299
	Fat	-0.115	-0.127
	Carbohydrate	0.300	0.201
	Sugar	0.461	0.293
Diabetes	Calorie	-0.077	-0.048
	Protein	-0.162	-0.022
	Fat	-0.123	-0.063
	Carbohydrate	0.173	0.106
	Sugar	0.142	0.066
Health Expend.	Calorie	0.098	0.110
	Protein	-0.083	-0.022
	Fat	0.197	0.141
	Carbohydrate	-0.064	-0.015
	Sugar	0.134	0.069

Conclusion

Conclusion

Results:

- Diversity of ingredients varies largely across the continents, mostly affected by migration trends.
- Strong similarities found between cuisines in neighboring countries.
- Clear correlation exists between obesity and cuisines that featured more sugar and carbohydrates.
- Protein-rich cuisines associates with fewer health problems.

Future Work:

- Recipe recommendation
- Modeling flavors with ingredients
- Ranking healthiest cuisines



Thank You!

Any Questions?

sajadmanesh@ce.sharif.edu g.stringhini@ucl.ac.uk @gianluca_string