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

April 5, 2017
Dataset
Dataset

- **Yummly**
  - Over 157K recipes in more than 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.

- **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*
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:
  \[ JS(P, Q) = \frac{1}{2} [KL(P \parallel M) + KL(Q \parallel M)] \]
  \[ M = \frac{1}{2} (P + Q) \]

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:
  \[ \frac{1}{2} [KL(P \parallel Q) + KL(Q \parallel P)] \]
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 cuisine denotes 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
Nutrients:

- Carbohydrate
- Calorie
- Fat
- Protein
- Sugar

Health Statistics:

- Diabetes
- Obesity
- Health Expenditure

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

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