Ingredients matching in bakery products

Tome Eftimov
Barbara Koroušić Seljak

{tome.eftimov, barbara.korousic}@ijs.si
Objectives

• Relevant and interesting relations between recipes’ ingredients

• Target data – bakery recipes
Finally settling down to my vegan, gluten free, soy free, antibiotics free, raw, non GMO, organic, fat free, low carb meal!
Introduction

• Data extraction

• Data preprocessing

• Association rules

• Visualization of the results
The knowledge discovery process
Data

- Collection of 1,900 bakery recipes written in English

- Data preprocessing: removed many adjectives associated with the cooking process, located synonyms

- Transformation: document-term matrix, transactional matrix
Basic statistics of the data set (1/2)

- Our transformed data contains 1,900 recipes and for each transaction we needed to consider the presence of 542 ingredients.

- The data set is rather sparse with a density just above 1.65%.

- The most popular ingredients are “salt”, “all-purpose flour”, “sugar” and “egg”.
Basic statistics of the data set (2/2)

- We rejected the 4 most frequently used ingredients for the analysis.

- Our data set contains 1,900 recipes and for each transaction we needed to consider the presence of 538 ingredients.

- The data set is rather sparse with density just above 1.13%.

- Average transaction contains less than 7 ingredients.
Methods

• Apriori algorithm

• Support, confidence, lift

• R programming language: “arules”, “arulesViz”
Evaluation

- Support = 0.005
- Confidence = 0.75
- The number of the discovered rules is 1,235
- After we pruned the redundant rules, the number of the discovered rules is 594
## The top 15 discovered rules

<table>
<thead>
<tr>
<th>#</th>
<th>LHS</th>
<th>RHS</th>
<th>support</th>
<th>confidence</th>
<th>lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{bread-flour, caraway-seed}</td>
<td>{rye-flour}</td>
<td>0.006</td>
<td>0.928</td>
<td>45.238</td>
</tr>
<tr>
<td>2</td>
<td>{active-yeast, caraway-seed}</td>
<td>{rye-flour}</td>
<td>0.008</td>
<td>0.888</td>
<td>43.304</td>
</tr>
<tr>
<td>3</td>
<td>{caraway-seed, water}</td>
<td>{rye-flour}</td>
<td>0.008</td>
<td>0.888</td>
<td>43.304</td>
</tr>
<tr>
<td>4</td>
<td>{cranberries, orange-juice}</td>
<td>{orange-zest}</td>
<td>0.005</td>
<td>0.846</td>
<td>30.333</td>
</tr>
<tr>
<td>5</td>
<td>{orange-juice, walnuts}</td>
<td>{orange-zest}</td>
<td>0.005</td>
<td>0.833</td>
<td>29.874</td>
</tr>
<tr>
<td>6</td>
<td>{baking-soda, cinnamon, molasses}</td>
<td>{ginger}</td>
<td>0.006</td>
<td>0.800</td>
<td>24.126</td>
</tr>
<tr>
<td>7</td>
<td>{garlic-powder, milk}</td>
<td>{cheddar-cheese}</td>
<td>0.005</td>
<td>0.769</td>
<td>21.181</td>
</tr>
<tr>
<td>8</td>
<td>{cream-cheese, milk, vanilla-extract}</td>
<td>{confectioners-sugar}</td>
<td>0.005</td>
<td>0.909</td>
<td>16.142</td>
</tr>
<tr>
<td>9</td>
<td>{baking-soda, cinnamon, nutmeg, water}</td>
<td>{pumpkin}</td>
<td>0.007</td>
<td>0.823</td>
<td>14.901</td>
</tr>
<tr>
<td>10</td>
<td>{baking-soda, nutmeg, water}</td>
<td>{pumpkin}</td>
<td>0.007</td>
<td>0.789</td>
<td>14.285</td>
</tr>
<tr>
<td>11</td>
<td>{butter, cream-cheese, milk}</td>
<td>{confectioners-sugar}</td>
<td>0.005</td>
<td>0.785</td>
<td>13.951</td>
</tr>
<tr>
<td>12</td>
<td>{cinnamon, pumpkin-pie}</td>
<td>{pumpkin}</td>
<td>0.005</td>
<td>0.769</td>
<td>13.919</td>
</tr>
<tr>
<td>13</td>
<td>{allspice, water}</td>
<td>{pumpkin}</td>
<td>0.005</td>
<td>0.769</td>
<td>13.919</td>
</tr>
<tr>
<td>14</td>
<td>{pumpkin-pie, vegetable-oil}</td>
<td>{pumpkin}</td>
<td>0.006</td>
<td>0.764</td>
<td>13.837</td>
</tr>
<tr>
<td>15</td>
<td>{bread-flour, butter, water, wheat-flour}</td>
<td>{honey}</td>
<td>0.005</td>
<td>0.833</td>
<td>10.021</td>
</tr>
</tbody>
</table>
Graph visualization
Parallel coordinates plot

Parallel coordinates plot for 30 rules

- rye-flour
- orange-zest
- ginger
- caraway-seed
- butter
- cheddar-cheese
- cranberries
- confectioners-sugar
- orange-juice
- molasses
- milk
- nutmeg
- brown-sugar
- allspice
- raisins
- cinnamon
- garlic-powder
- pumpkin
- baking-soda
- vegetable-oil
- active-yeast
- wheat-flour
- cloves
- bread-machine-yeast
- cream-cheese
- vanilla-extract
- water
- walnuts
- bread-flour
- oats
- pumpkin-pie
- honey
- flax-seed
Grouped matrix visualization

Grouped matrix for 594 rules
Conclusion (1/3)

- Expected rules:
  - \{yeast\} $\rightarrow$ \{water\}
  - \{apple\} $\rightarrow$ \{cinnamon\}
Conclusion (2/3)

• Unexpected rules:
  • \{baking soda, cinnamon, molasses\} → \{ginger\}
  • \{baking soda, nutmeg, water\} → \{pumpkin\}
Conclusion (3/3)

- This analysis allows us to see how the ingredients are combined in bakery recipes.

- The information is very important for food compilers who need to collect analytical data for food items frequently used in national dietary surveys based on foods and recipes.
Further work

• To analyze these combinations in order to determine the nutritional properties for different values of quantity-unit pair for each ingredient

• Compare these relations with the relations provided by Foodparing
References

• Data Source: http://allrecipes.com/Recipes/Bread/Main.aspx?prop24=hn_browsedeep&evt19=1


