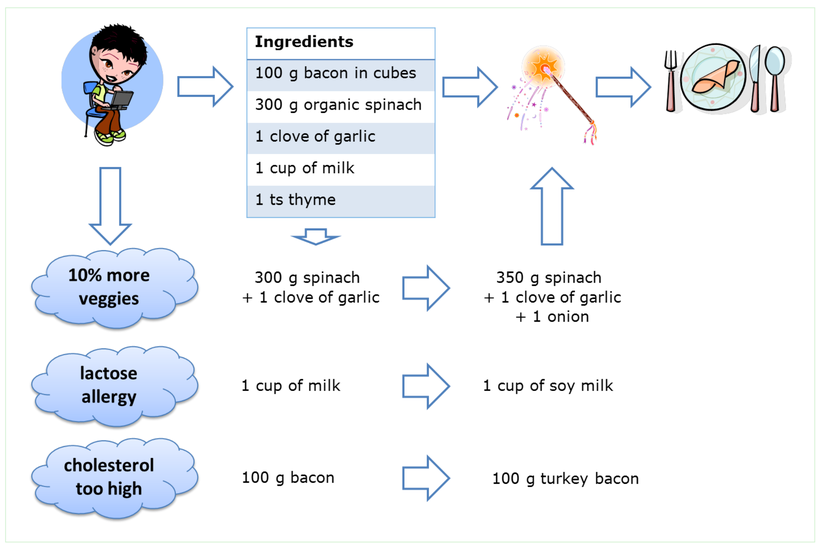
Automatic Extraction of Ingredient's Substitutes

Corrado Boscarino [1], Nicole J.J.P. Koenderink [2], Vladimir Nedovic [3], Jan L. Top [1,2]
[1] VU University Amsterdam, [2] Wageningen UR Food & Biobased Research, [3] Flavourspace Amsterdam
The Netherlands

Introduction

Our dream: **automatically customise recipes to personal profiles**. We want to take into account biological (allergy, nutritional needs) and psychological (like/dislike ingredients, personal intentions) factors, but also situational (occasion, meal companions) and social (religion, cultural norms) aspects. Also, extrinsic food properties (sustainable, organic) are relevant.

To do this, we need to know **what are appropriate suggestions given both the personal profile and the recipe involved.**



Approach

Question: can we automatically find ingredient substitutions that fit the recipe context?

Needed resources:
- Expert knowledge captured in substitution rules
- Context information from statistics of large recipe collections
- Food ingredient knowledge captured in an ontology

Expert Substitution Rules

We use Cook's Thesaurus (http://www.foodsubs.com) as repository (CT) of substitution rules.

Example: listed substiutes for **bacon:**
turkey bacon (less fat, doesn't shrink as much as regular bacon) OR *vegetarian bacon* OR *imitation bacon bits* OR *smoked ham* OR *Canadian bacon (leaner)* OR *pancetta* OR *salt pork (in sauces)* OR *smoked sausage* OR *ham* OR *fatback (for barding)*

Hypothesis: the substitution rules contain **implicit knowledge about the substitution context**. They expect that the readers have some background knowledge to assess the rules.

Statistics of Recipe Collections

We have used a repository (AH) of 12,515 recipes from the Dutch website Albert Heijn (http://www.ah.nl/allerhande/recepten).

Assumptions:
- the AH set is sufficiently large to represent typical recipes
- we can show a proof of principle of our method by just looking at the presence of ingredients, disregarding quantities and preparation methods.

We have
- reduced each recipe to its set of ingredients
- manually simplified the ingredients to create a uniformly specified list
  - remove taste-irrelevant adjectives
  - enforce uniform spelling
  - keep only plural or singular form
- reduced the list by removing ingredients that occur only once in the recipe collection

This results in a 2199-dimensional ingredient space (due to the presence of 2199 unique ingredients).

We apply **Latent Dirichlet Allocation** to limit the complexity of the space. Each recipe is represented as a point in the ingredient space. Neighbouring recipes are generalised to a **topic**. Any recipe can be represented as linear combination of topics.

Food Ingredient Ontology

We have taken the category structure in Cook's Thesaurus and modelled is as an ontology. We have created a mapping between the 2199 AH ingredients onto this CT ontology. It contains categories of ingredients, such as "oils" in which all types of oils – olive oil, sesame oil, sunflower oil,... – are represented.

Results

We have looked at the task of **finding substitutes in selected recipes and ranking the candidate substitutes**. We have chosen a reduced recipe space spanned by 50 topics. Next, we have performed a **qualitative analysis** of the method.

Result 1: Qualitatively, LDA topics correspond to ingredients sets typical for a particular dish type, e.g. a cake or a pasta. Examples of topics in the AH dataset are:
- *TOPIC 1: eggs 0.0868, butter 0.05, sugar 0.05, whipped cream 0.04, flour 0.03, casting sugar 0.02*
- *TOPIC 2: olive oil 0.09, onion 0.07, garlic 0.05, bell pepper 0.02, tomato dice 0.02, tomato 0.02*

Result 2: Topic densities of document's mixtures and their neighbours can be used for ranking candidate substitutes from expert rules. Example:
- *Ciabatta pizza = olive oil, vegetable mix with egg plant, ciabatta, sun dried tomato sauce, tuna, mozzarella*
- *Olive oil substitutes: walnut or almond or hazelnut oil (for cold dishes and salads)* OR *corn or peanut oil (for sautéing)* OR *cooking spray (for greasing pans)*
- *Olive oil is a type of Oil & Spray. Olive oil possibly substitutes all instances of the class Oil & Spray.*

The representation of the recipe Ciabatta pizza in topic space is: $DP(1647,:) = 1x\ T_{37} + 5x\ T_{48}$
- *Topic 37 = 0.13 basil, 0.09 mozzarella, 0.08 cherry tomato, 0.06 olive oil, 0.06 olive oil extra virgin, 0.05 rocket salad*
- *Topic 48 = 0.13 olive oil, 0.11 capers, 0.11 garlic, 0.10 black olives, 0.09 anchovies, 0.05 lemon*

Extra virgin olive oil is the highest ranking substitution suggestion, a type of Oil & Spray. Not unexpected, due to its use in Italian cuisine.

Conclusions

- After qualitative analysis, we conclude that substitution rules can be found by looking at ingredients based on their position in topic space, in the ontology and in explicit substitution rules.
- Food ontologies seem to be useful for finding similar types of ingredients; topic space for finding similar types of dishes.

Acknowledgements

This publication was supported by the Dutch national program COMMIT and by the Dutch TopSector program KV 1309 019.

VU University of Amsterdam
De Boelelaan 1081, 1081 HV Amsterdam
Contact: Jan.Top@wur.nl,

Food & Biobased Research – Information Management
P.O. Box 17, 6700 AB Wageningen
Contact: Nicole.Koenderink@wur.nl,
http://url.fbresearch.nl/IM

Flavourspace
Mauritskade 55C, 1092 AD Amsterdam
Contact: v@flavourspace.com