Donnerstag
09.03.2017
12:45 – 13:45
B4 1, Foyer

Distributional and Neural Gastronomics

Laura Bostan
Trento University
sarnthil@gmail.com

Jonathan Oberländer
Trento University
jonathan.oberlaender@gmail.com

The goal of this project is to explore the potential for computational creativity in gastronomy by developing a data-driven recipe generation system. The completed system should be able to learn events from a corpus of recipes, learn which events follow which by looking at ordered event co-occurrences, and then sequence events to form the basis of a new recipe using a model trained on event sequences of previous recipes.

We used Open Recipes[1], a database of recipes, automatically collected by crawler scripts.This dataset was enriched with cooking instructions by crawling the original websites of the recipes. After getting a collection of cleaned recipes; ingredients, modifiers, units, utensils and events were extracted. Cleaned ingredients were hard to obtain, because even though they are specifically listed per recipe normalizing them was difficult because ingredient names are intertwined with amounts, units, and other noise. Events of cooking are simply verbs that take an ingredient as an argument. This requires previous POS-tagging and parsing of the cooking instructions. We noted that the approach used by Kiddon et al. (2015) is more elaborate and we plan to follow their heuristics for extracting events.

We then created a semantic vector space using co-occurrence counts (ingredients co-occurring with ingredients, modifiers, amounts, and units), reweighted the resulting matrix with PPMI and reduced the dimensions to 100 using SVD. These enriched learned representations are used as the inputs to our generator. We experimented with building our generator using different neural models (char/word based LSTM language models, seq2seq) on the possible events given the ingredients.

We have thought of different ways of evaluating the quality of the automatically generated recipes, such as manual linguistic analysis of the output recipe, blind taste tests, or Turing tests comparing "real" recipes with generated ones and with competitor systems.

References: • Kiddon C., et al. (2015): Mise en Place: Unsupervised Interpretation of Instructional

[1] https://github.com/fictivekin/openrecipes

Recipes. Proceedings of the 2015 EMNLP.