

Identifying and Combining Ingredients from Natural Instructions

Allen Park

Instructables Data

I scraped instructables from [instructables.com](https://www.instructables.com). Each set of instructions was made of one or more steps.

Example steps:

“Drill a quarter inch hole in the bottle cap. Insert the quarter inch tubing connector (you can buy one at Home Depot for 99 cents) into the hole and glue using glue gun on both sides.”

“The ingredients include seasonal winter fruits, which makes this a perfect holiday dessert. It's absolutely delicious and will surely be a favorite at any party, plus everyone will be able to enjoy it since it doesn't include any dairy or gluten.”

“Place the straws as far apart as possible for stability, and perpendicular to the bottle center-line.”

Ingredient Data

I obtained lists of ingredients from online sources. The end result was a list of 2152 words.

Example ingredients:

“Beef Flank Steak Choice Ln 0"Tr-Brsd”

“Eggplant Cubes Cooked-Drnd”

“Fresh Lemon Juice”

<http://www.esha.com/solutions/additional-databases/>

<http://tna.europarchive.org/20110116113217/http://www.food.gov.uk/science/dietarysurveys/dietsurveys/>

Unsupervised Training of Ingredient Values

Goal: give every word an ingredient weight

Idea: Use word co-occurrences with other words in steps

1. Initialize weights based on nouns in prior ingredient list.
2. For every word:
3. For every step the word is in, compute the average weight of the step.
4. Compute the average step weight.
5. Average the average step weight and the previous word weight to get the new word weight.
6. Repeat (go to step 2)

Food Recipe Classification

Goal: Classify Instructables into food recipe or not

6830 instructables	Food words in titles ≥ 2 Baseline	Food fraction in titles ≥ 0.34 Baseline	Food fraction in steps ≥ 0.12 Baseline	Ingredient unigram weights 0.061 Unsupervised	Ingredient bigram weights 0.064 Unsupervised
Accuracy	0.8730	0.8347	0.9685	0.9836	0.9602
Precision	0.6528	0.9497	0.8958	0.9482	0.8580
Recall	0.6800	0.1187	0.9402	0.9646	0.9427
F1	0.6661	0.2110	0.9175	0.9563	0.8984

Ingredient Step Supervised Classification

Problem: Many Instructables have a step where they list ingredients

Goal: Classify steps on whether they contain a list or not

429 steps	List tag presence Baseline	Number fraction <= 0.05 Baseline	Noun fraction <= 0.36 Baseline	Verb fraction >= 0.11 Baseline	Multi-layer perceptron Supervised	Simple classifier >= 0.5 Unsupervised
Accuracy	0.8997	0.8671	0.8787	0.8414	0.9457	0.8764
Precision	0.6363	0.4230	0.4666	0.3855	0.64133	0.4655
Recall	0.2857	0.4489	0.4285	0.6530	0.6154	0.5510
F1	0.3943	0.4356	0.4468	0.4848	0.6281	0.5046

Ingredient Word Unsupervised Classification

Goal: Classify words by whether they're an ingredient or not

1112 words	Food word Baseline	Food noun Baseline	Simple classifier \geq 0.97 Unsupervised	Multi-layer perceptron Unsupervised	Multi-layer perceptron Supervised
Accuracy	0.7913	0.8812	0.8862	0.8810	0.9004
Precision	0.5	0.7451	0.7313	0.6623	0.7254
Recall	0.7543	0.6681	0.7101	0.8879	0.8222
F1	0.6013	0.7045	0.7205	0.7587	0.7708

Ingredient Word Unsupervised Classification

For training data, label nouns with their ingredient values with a cutoff value of 0.07. All other parts of speech were not ingredients.

The features were a trigram of 50-dimensional GLOVE word vectors, a variety of booleans like is alpha, is ascii, is lowercase, etc., and a 1-hot vector based on the part of speech.

Basic Action Parsing

Goal: Parse actions from a instructable with labeled ingredients

“First cut your onions into small pieces, then cut your garlic cloves into very fine cubes. Put some oil in your pot and roast the onions and garlic until golden-brown. Take the pot from the stove.”

Verb	Text	Labeled ingredients
cut	your onions	onions
cut	into small pieces	pieces
cut	your garlic cloves	garlic, cloves
cut	into very fine cubes	cubes
Put	some oil	oil
Put	in your pot	pot
	and so on	