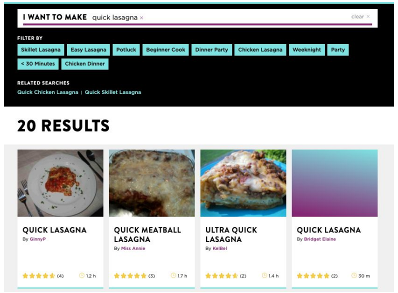
A Recipe for Success

INFO 254 Final Project
Fall 2020
Samantha Carr, Lia Chin-Purcell, Siqi Wu, Kailin Koch

Motivation

Despite the volume of recipes are available online, **common recipe queries are challenging.**

Current options given are too narrow and don't manage constraints well.



I WANT TO MAKE quick lasagna ×

20 RESULTS

QUICK LASAGNA By GinnyP (4) 1.2 h

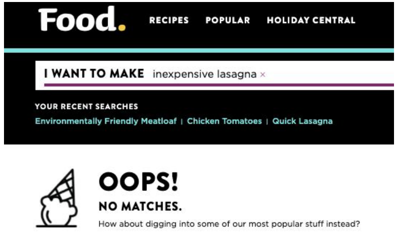
QUICK MEATBALL LASAGNA By Miss Annie (3) 1.7 h

ULTRA QUICK LASAGNA By KelBel (2) 1.4 h

QUICK LASAGNA By Bridget Elaine (2) 30 m

"quick lasagna weeknight"

Too literal - it's only giving us lasagna recipes, not recipes that include the same basic ingredients and seasonings that may suit the purpose.



Food.
RECIPES
POPULAR
HOLIDAY CENTRAL

I WANT TO MAKE inexpensive lasagna ×

YOUR RECENT SEARCHES
Environmentally Friendly Meatloaf | Chicken Tomatoes | Quick Lasagna

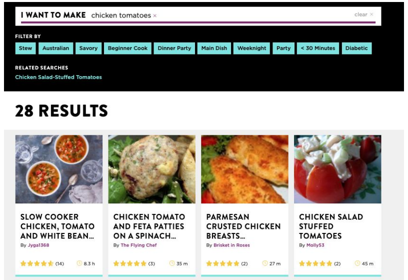
OOPS!

NO MATCHES.

How about digging into some of our most popular stuff instead?

"inexpensive lasagna"

Few options - there's not an easy way to find basic adjustments of recipes like this.



I WANT TO MAKE chicken tomatoes ×

28 RESULTS

SLOW COOKER CHICKEN, TOMATO AND WHITE BEAN… By Jyga1368 (14) 8.3 h

CHICKEN TOMATO AND FETA PATTIES ON A SPINACH… By The Flying Chef (3) 35 m

PARMESAN CRUSTED CHICKEN BREASTS… By Brisket in Roses (2) 27 m

CHICKEN SALAD STUFFED TOMATOES By Molly53 (2) 45 m

"chicken tomatoes"

Too many ingredients - while you can find recipes that include ingredients you have, there's no easy way to build a recipe based only on the ingredients you have.

Audience

Home cooks who want to find new recipes and ingredient substitutions to cook the food they want while still meeting their specific preferences or needs.

Research Goals

Skip-gram

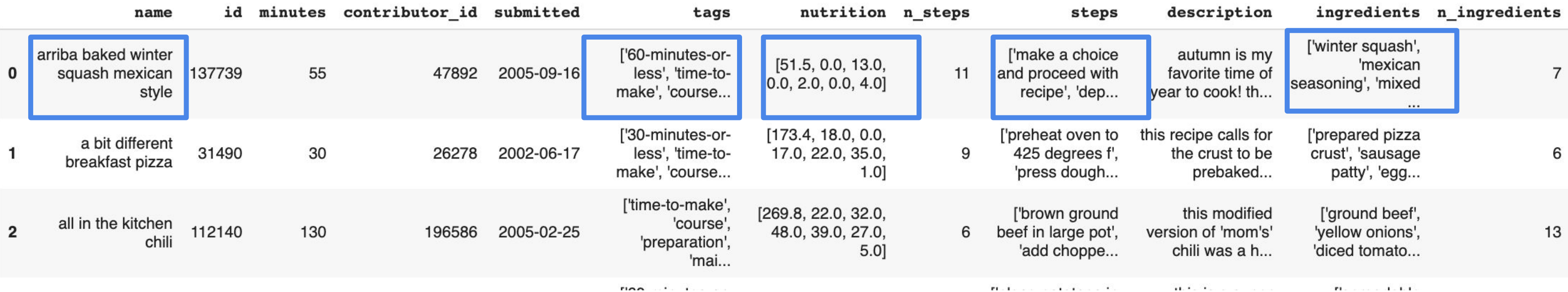
- Provide alternatives to dishes that satisfy different constraints
  - Faster to make
  - Good for kids
  - Possible with ingredients you already have
  - Similar to a dish you’ve already made, but new and adventurous
- Suggesting ingredient alternatives that are:
  - Cheaper
  - Environmentally friendly
  - Healthier

RNN

- Generate new recipes possibilities from existing recipe corpus, given starting steps
- Generate new recipes based on ingredients, time or occasion

Dataset

- Dataset on Kaggle, data from Food.com
- Over 230,000 recipes and 700,000 recipe reviews covering 18 years of user interactions and uploads
- 12 Features - We focused on recipe name, recipe steps, ingredients, tags and nutrition.
- Data cleaning and preparation
  - Preparing ingredients for the model - What is an ingredient?
  - Incorporating non-ingredients into recipes, such as recipie name and tags



	name	id	minutes	contributor_id	submitted	tags	nutrition	n_steps	steps	description	ingredients	n_ingredients
0	arriba baked winter squash mexican style	137739	55	47892	2005-09-16	['60-minutes-or-less', 'time-to-make', 'course...	[51.5, 0.0, 13.0, 0.0, 2.0, 0.0, 4.0]	11	['make a choice and proceed with recipe', 'dep...	autumn is my favorite time of year to cook! th...	['winter squash', 'mexican seasoning', 'mixed ...	7
1	a bit different breakfast pizza	31490	30	26278	2002-06-17	['30-minutes-or-less', 'time-to-make', 'course...	[173.4, 18.0, 0.0, 17.0, 22.0, 35.0, 1.0]	9	['preheat oven to 425 degrees f', 'press dough...	this recipe calls for the crust to be prebaked...	['prepared pizza crust', 'sausage patty', 'egg...	6
2	all in the kitchen chili	112140	130	196586	2005-02-25	['time-to-make', 'course', 'preparation', 'mai...	[269.8, 22.0, 32.0, 48.0, 39.0, 27.0, 5.0]	6	['brown ground beef in large pot', 'add choppe...	this modified version of 'mom's' chili was a h...	['ground beef', 'yellow onions', 'diced tomato...	13

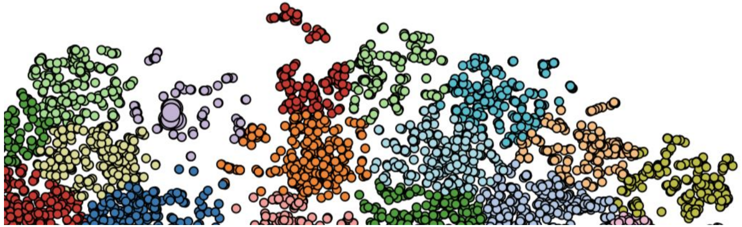
Approaches

- Word2Vec - Gensim model
    - Data Engineering for certain parameters like “Environmentally Friendly” ingredients, transforming tags
    - Dimensionality reduction using TSNE, visualization of points
    - Clustering using k-means, use of elbow method to find an optimal k
    - Evaluated compared to google news and using ground truth distance evaluations
- Word-Based RNN
    - Similar to model used in class, but instead of using characters as the timestep, using words
    - Allow generation of sequences of words

Results

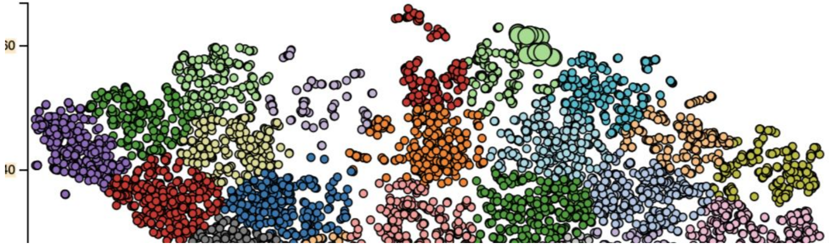
	Meal				Ingredients		
	Fast	Good for Kids	Ingredients Based	New Alternatives	Healthier	Environmentally Friendly	Cheaper
	Steak + fast + dinner - slow	*Pasta + dinner + kids*	*Burger - meat*	*Lasagna + Pasta - tomatoes*	*Ground Beef - Unhealthy*	*Olive Oil - emissions*	*Steak - Expensive*
Our Word2Vec Models	hamburgers	Cheese tortellini	wrap	tetrazzini	Ground turkey	Canola Oil	pork
Google News Word2Vec	brunch	supper	Cheese burger	Chicken cacciatore	Beef imports	pomegranate molasses	ribeye
Top Google Search Result	Recipe: Garlic butter steak bites	Article: Kids' pasta recipes	Recipe: The best Veggie Burger	Recipe: Heartburn friendly tomato sauce	Article: Healthy meat substitutes	Article: 5 delicious, sustainable olive oil brands	Article: 4 Steaks You Can Actually Afford

Interesting Findings - Clustering



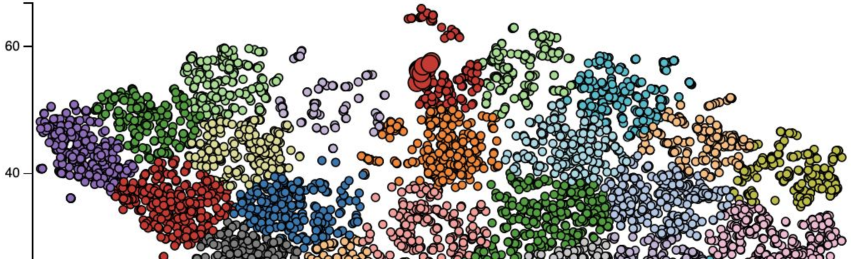
word
apricot_jam: 1
apricot_preserves: 1
bourbon: 1
cherry_preserves: 1
apple_jelly: 1

cluster
27: 11



word
anaheim_chili: 1
anaheim_chilies: 1
chipotle_chile: 1
chipotle_chile_in_adobo: 1
adobo_sauce: 1

cluster
45: 46



word
chiligarlic_sauce: 1
chinese_five_spice_powder: 1
dark_sesame_oil: 1
dark_soy_sauce: 1
chili_paste: 1

cluster
52: 27

Validation of Findings

	Easy			Medium		Tough	
	Cosine similarity - “Chicken” to “Turkey” > “Chicken” to “Coffee”	Most Similar to Pasta - types of pasta	Doesn’t Match (‘chicken’, ‘potato’, ‘cookie’)	“Burger” - “Meat” ~ ~ sandwich	“Turkey broth” more similar to “chicken broth” than “turkey”	“Olive oil” + “environmentally friendly” = another type of oil	“Spare ribs” + “environmentally friendly” = meat alternative
Google news	0.62825197 0.2746489	Pastas, tomato_sauce, polenta	cookie	cheeseburger	NA	extra_virgin_olive	beef_fajitas
Model 1	0.4888747 -0.41626862	Penne, rigatoni, spaghetti	cookie	Mixed french herbs	Turkey - 0.5364784 Chicken - 0.6643304	Canola oil	Vegetarian hot dogs
Model2 Ingredients names, top tags, recipe name	0.6654525 -0.27789456	Penne pasta,spaghetti, angel hair pasta	cookie	casserole	Turkey - 0.52861464 Chicken broth - 0.64949924	NA	NA
Model 3 Ingredients, names, tags freq >=5, window=15	0.6251654 -0.35022154	Tortellini, penne, bow tie pasta	cookie	wraps	Chicken broth 0.5410633 Turkey 0.43914166	NA	NA

Validation of Findings - Under constraints

	Tough		
	“Desserts” - “Butter”	“Cookies” + “Vegan” - “Dairy”	“Dinner” + “Kids”
Google news	Desserts, appetizer, brunch	Cupcakes, brownies	Dinners, brunch, supper
Model2 Ingredients names, top tags, recipe name	Velvet martini, berry protein smoothie	Banana oat bars, dried apple rings	Pizza snack cups, fat meatballs
Model 3 freq>5 ingredients, names and basic tags Window = 15	Cranberry jello	Scones, muffins, biscotti	Margarine Mashed potatoes

Interesting Findings - RNN

Given seed text of a full known sequence

“muffins need only minutes with this recipe, i usually always make muffins to cut down bake time, preheat oven, to combine cookie crumbs butter press firmly on bottom of inch springform pan, in large mixing bowl beat cream cheese until fluffy, beat in sweetened condensed milk eggs vanilla in small **//** bowl and blend together, add well mashed bananas in a bowl, youll until well blended, crush flour and mix in butter, add eggs and vanilla to mixer, blend well, add chips and extracts, beating until firm peaks form, serve with rice and lb of ground meat, put dumplings on top”

Given seed text of beginning of known sequence

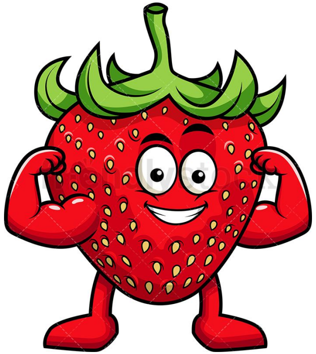
“with an electric mixer or wire whisk beat eggs until frothy whisk in flour **//** butter and next soda add in tomatoes to horseradish and lime zest and mix well add shrimp and flour whisk together well add liquid to the cornstarch until well blended add eggs garlic and butter until smooth add eggs and mix add oil and stir until smooth and coated add the bananas broth celery seed garlic extract until well blended”

Given seed text of non-trained sequence

“whisk eggs together and add cheese **//** and add pinches of salt substitute over low oil in a large bowl mix together and cup dry ingredients pour over dough bake for minutes or until pancake is puffed”

Challenges We Met

1. Big dataset and limited computing power
2. Bringing in other data sets for ingredient environmental footprint and ingredient cost
3. How to define ingredients (“baking_soda” vs “baking” and “soda”)
4. How to parse descriptors
5. Distinguishing between meals, ingredients, and descriptors



Trials and Errors

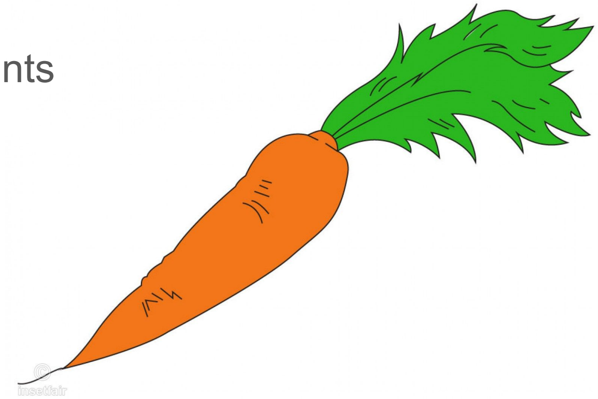
- A character-Level LSTM RNN
- Using whole data-set to train the RNN model
- Breaking up the ingredients when training
    a. Models learn “winter squash” as “winter” and “squash”
- Calculating calories for each ingredient based on total calories of recipe
    a. General problems with ideas of healthiness, especially when basing off of just calories
- Trying to generate a recipe from a trained RNN model based on ingredient inputs only
- Creating a validation set for evaluation

Results and Conclusions

- Overall, our Word2Vec model does a decent job of offering alternatives based on information included in the dataset. For example, our model can give an alternative for kid-friendly dishes (example) or healthier alternatives.
- Our Word2Vec model does a worse job of offering alternatives for information brought in from outside sources, also may be a limitation of recipe writing (recipe writers may consider the overall healthiness of a given recipes, but it's less likely they would consider the environmental impact). One way to solve this is spend time data cleaning/feature engineering so that datasets can be combined with more information on environmental impact and cost.
- Our RNN model was able to output logical steps and follow recipe sentence patterns, and if given enough starting seed text, could incorporate existing ingredients mentioned into following steps generated
- Because the RNN model was trained on recipe instructions, ingredient-only seed text wasn’t able to generate cohesive or relevant instructions

In the future...

- App that allows users to create a “food profile” that is then used by the app for recipe suggestions and generation
    - Meals they enjoy
    - Dietary preferences
    - Considerations (Environmentally-friendly, Vegetarian, Reduced-sodium)
- Users can pull up the app for times when they want to try something new, or get suggestions for an alternative option given their preferences
- However, further work is needed to ensure:
    - There was better differentiation between dishes and ingredients
    - We had more computing power
    - Model could learn from regularly updated corpus



bloopers

- vanilla_ice_cream - unhealthy = ice_cubes
- burger - meat = pickle_juice
- dinner + healthy = kids