Kevin Manella - 829549
Stefano Talamona - 822452

Food Recommender System

Abstract

The project is about the recommendation of food recipes to the user. For example, when a person browses a cooking website he/she may be interested to a similar recipe to the one he/she is looking at.

In order to accomplish this task it's better to use different types of data, such as images and texts.

We propose a **food recommender system** based on **multimodal embedding** that is able to suggest a recipe similar to the one given in input.

Food Recommendation: Framework, Existing Solutions and Challenges

W. Min, S. Jiang, R. Jain - 2020

- State of the art for food recommendation systems
- Multimodal data is a key aspect of food representation

Is the suggested food your desired?: Multi-modal recipe recommendation with demand-based knowledge graph

Z. Lei, A. Ul Haq, A. Zeb, M. Suzauddola, D. Zhang - 2021

- Feature extraction from food images with a deep CNN
- Embedding of textual data with Doc2Vec

Recipe Recognition with Large Multimodal Food Dataset

X. Wang, D. Kumar, N. Thome, M. Cord, F. Precioso - 2015

- Importance of dataset quality and dimensionality
- Choice of the dimension of the image embedding vector

Dataset

- *Food Ingredients and Recipes Dataset with Images* from Kaggle
- About 13500 food recipes with corresponding images (no exact match between rows of the dataset and images → data cleaning)
- Content:
  - ***Title***
  - ***Ingredients***
  - *Instructions*
  - ***Image_Name***
  - *Cleaned_Ingredients*

Remove unwanted characters

- Applied to ***Title*** and ***Ingredients***
- Lowercase conversion
- Keeping only letters and spaces

['1 (3½–4-lb.) whole chicken', '2¾ tsp. kosher salt, divided, plus more']

lb whole chicken tsp kosher salt divided plus more

Remove unwanted words

- Applied to ***Ingredients***
- Removing non-food words (“tablespoon”, “sliced”, “and”, etc...)

lb whole chicken tsp kosher salt divided plus more

chicken kosher salt

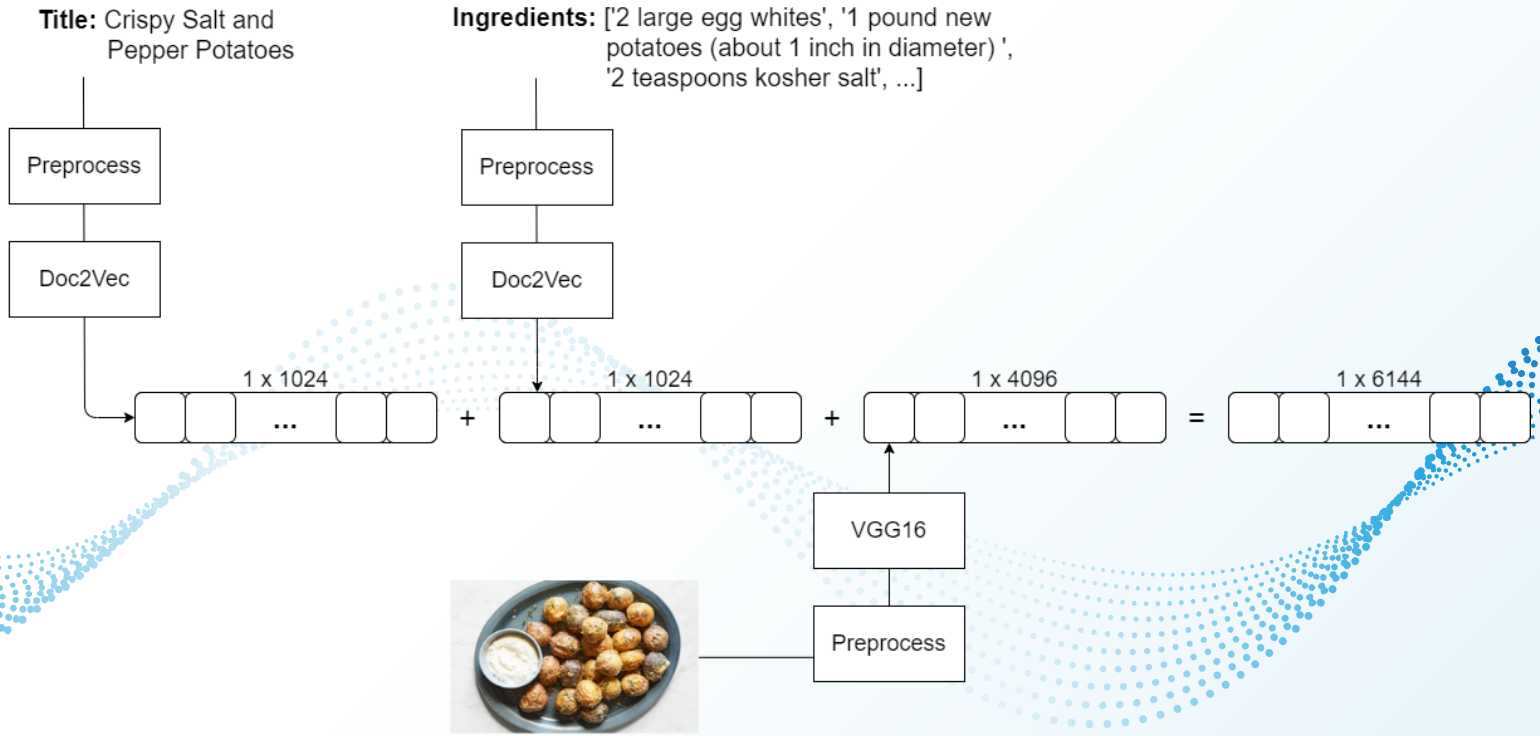
Document Embedding

- ***Doc2Vec*** from ***Gensim*** library
- Two models: one for ***Title*** and one for ***Ingredients***
- Training parameters:
  - *vector_size = 1024*
  - *epochs = 30*

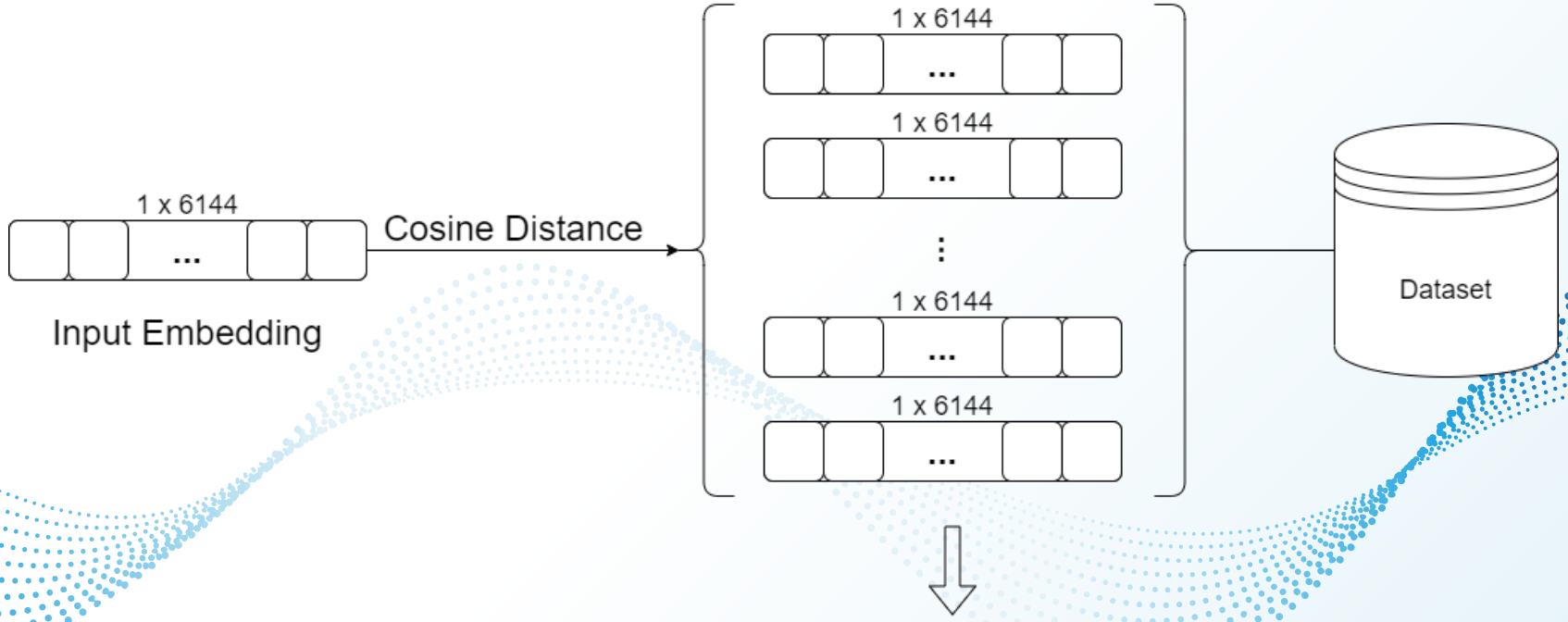
Image Embedding

- ***VGG16*** from ***TensorFlow*** library
- Pre-trained on *ImageNet*
- Cut point: *fc1* layer
- Input size: (224, 224, 3)
- Output size: (1, 4096)

Multimodal Embedding



Cosine Distance



Discussion of the Results

Images and Texts both similar

Existing Data

Input

Title: neapolitan style pizza pizza alla napoletana

Ingredients: water envelope active dry yeast extra virgin olive oil needed flour needed kosher salt buffalo mozzarella san marzano tomatoes basil leaves dried sicilian oregano black pepper

Output

Title: cherry tomato pizza margherita

Ingredients: refrigerated pizza dough extra virgin olive oil cherry tomatoes garlic clove pressed fennel seeds plastic dried red pepper mozzarella water ovoline milk mozzarella basil leaves leaves

Images and Texts both similar

New Data

Input

Title: classic hamburger

Ingredients: onion oil ketchup mayonnaise white vinegar ground beef ground pepper bread tomato lettuce

Output

Title: the ultimate hamburger

Ingredients: chuck sirloin salt kosher sea black pepper black pepper size onion sandwich bread butter size gorgeous luscious ripe red tomatoes boston lettuce leaves iceberg lettuce cheese sauce

Similar Images, Texts not

Existing Data

Input

Title: green curry vinegar chicken

Ingredients: shallots green beans peppadew peppers brine white wine vinegar extra virgin olive oil thai green curry paste chicken thighs drumsticks kosher salt

Output

Title: parmesan bread pudding with broccoli rabe and pancetta

Ingredients: olive oil garlic cloves red pepper flakes broccoli rabe rapini kosher salt black pepper eggs milk country white bread parmesan pancetta bacon

Similar Images, Texts not

New Data

Input

Title: lasagna stuffed shells

Ingredients: beef black pepper olive oil onion tomato paste dried oregano tomato sauce jumbo pasta shells cloves garlic whole milk ricotta cheese mozzarella cheese spinach basil lemon parmesan

Output

Title: simple is best dressing

Ingredients: unsalted butter baking day white bread yellow onions celery leaf parsley sage rosemary thyme kosher salt black pepper chicken broth vegetable broth eggs

Similar Texts, Images not

Existing Data

Input

Title: shaved zucchini salad with parmesan pine nuts

Ingredients: extra virgin olive oil lemon juice kosher salt black pepper dried red pepper zucchini basil pine nuts wedge parmesan cheese

Output

Title: grilled zucchini and leeks with walnuts and herbs

Ingredients: walnuts garlic clove lemon juice olive oil kosher salt black pepper leeks white pale green root attached zucchini leaf parsley leaves

Images and Texts not similar

Existing Data

Input

Title: tunisian vegetable salsa

Ingredients: garlic cloves plum tomatoes onions red bell peppers poblano chile japanese eggplant extra virgin olive oil lemon juice kosher salt black pepper dark green poblano chiles

Output

Title: lamb chili with masa harina dumplings

Ingredients: dried mild new mexico chiles water lamb shoulder black pepper salt lard vegetable oil onion garlic cloves turkish bay leaves california cumin dried oregano canned chipotle chiles adobo masa harina corn tortilla mix flour baking powder baking soda salt lard unsalted butter shaken buttermilk cilantro

Images and Texts not similar

New Data

Input

Title: spaghetti with meatballs

Ingredients: spaghetti ground beef bread crumbs parsley parmesan egg garlic cloves kosher salt red pepper virgin olive oil onion tomatoes bay ground black pepper

Output

Title: grilled shrimp and scallions with southeast asian dipping sauces

Ingredients: asian fish sauce lime juice sugar jalapeño chile rings soy sauce rice wine vinegar ginger cilantro coriander seeds colossal under per jumbo per shrimp tails intact deveined scallions white green vegetable oil kosher salt black pepper called nuoc mam nam pla asian asian section water

Conclusions

The model performs pretty well on some inputs, while it performs poorly on some others.

The performance of the model strongly depends on the quality of the dataset and of the input data. Many instances of the dataset contain dirty data, such as images where the food is not centered or not present at all and texts where there are many non-food words (it's difficult and time-consuming to remove all of them).

Future Works

- There isn't a quantitative metric for this task, so it could be useful to validate the model by collecting users feedback
- Try other DCNN architectures (ResNet, MobileNet, ...) or cut points of VGG16 to perform image embedding
- Try other approaches to perform document embedding (SentenceBERT, InferSent, UniversalSentenceEncoder, ...)
- Perform a further cleaning of the dataset or search for one with cleaner data
- Try a more sophisticated approach to join texts embeddings and images embeddings

Thanks

Any Questions?