Artificial Intelligence
AA 2018/2019

Analogy-based Recommender system of multi-ethnic dishes

Francesco Morano 794324

Stefano Pellicciolì 793414

Umberto Azadi 794276

Rationale of the work

Recently, even in the smallest town of Italy can be found an increasing number of **restaurants that offer dishes originated from all over the world**. With such a heterogeneous offer it is likely that the average person will not know a priori which type of cuisine or more specifically which dishes will be likely to meet his/her taste.

The **purpose of this project** is to support the customers in this specific decision making step, by **suggesting them a dish of a cuisine never tasted before** based on their preferences in a known cuisine, as for example the cuisine of their home country.

Main objectives

- Exploit a **multi-modal embedding** representation of dishes in order to accomplish the **analogical inference** through which the recommendation is obtained:
  - How to embed the recipes?
  - How to embed the images?
  - How to combine the two kinds of embedded features?
  - How can analogical inference be used as a recommender system?

- Work in an **open world assumption**:
  - Allow the users to add their own recipes and use them to obtain the recommendations;

- Keep into consideration **specific diets** such as: vegetarian, vegan, dairy-free, ...

Related Works

Experimenting Analogical Reasoning in Recommendation

(Hug et al. [1] 2015)

- **Objective:**
  - investigate the possibility of using analogy as the main underlying principle for implementing a prediction algorithm of the collaborative filtering type.

- The **analogical jump** is an unsound inference principle postulating that, given 4 vectors a, b, c, d such that the proportion holds on some components, then it should also hold on the remaining ones.

Algorithm 1. Analogy

Input: A set of known ratings R, a user u, an item i such that $r_{ui} \notin R$.
Output: $\hat{r}_{ui}$, an estimation of r_{ui}
Init:
$C = \varnothing$ // list of candidate ratings
for all users a, b, c such that

1. $r_{ai} \in R, r_{bi} \in R, r_{ci} \in R$
2. $r_{ai} - r_{bi} = r_{ci} - x$ is solvable // i.e. the solution $x \in [1, 5]$
3. $||(a - b) - (c - d)|| \leq \lambda$ // Analogy almost stands between a, b, c, d considered as real vectors

do
  $x \leftarrow r_{ci} - r_{ai} + r_{bi}$
  $C \leftarrow C \cup \{x\}$ // add x as a candidate rating
end for
$\hat{r}_{ui} = \underset{x \in C}{\text{aggr}}\, x$

$$\frac{\forall j \in J, a_j : b_j :: c_j : d_j}{\forall i \in [1, n] \setminus J, a_i : b_i :: c_i : d_i} \quad (analogical\ inference)$$

Recommender System based on Argumentation by Analogy

(Budán et al. [2] 2014)

- **Objective:**
  - Use **Defeasible Logic Programming** to increase the justifications and foundations that support a particular recommendation, by an **analogy process**.
- There exists a set of shared properties that permits projecting the conclusion, which is not derived systematically from the premises, but it constitutes a plausible recommendation, in the following way:

$$(P(S) \wedge Q(S)) \wedge (P(T) \Rightarrow Q(T))$$

Example 2 *Let A and B be two films where:*

$$\frac{\begin{array}{c}\texttt{Rating}(A, High) \wedge \texttt{Rating}(B, High) \\ \texttt{Genre}(A, Comedy) \wedge \texttt{Genre}(B, Comedy) \\ \texttt{Tags}(A, Excellent) \wedge \texttt{Tags}(B, Excellent) \\ \texttt{Recommendable}(A)\end{array}}{\texttt{Recommendable}(B)}$$

Knowing the values of `Rating`, `Genre`, *and* `Tags` *of films A and B, and if the first is recommendable, it is possible to determine if film B is recommendable. Then P (set of shared properties) is said to determine Q (the conclusion). In symbols: $P \succ Q$*

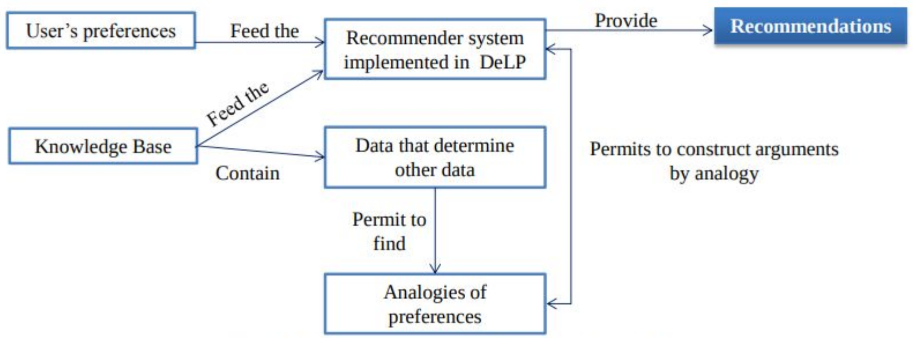


Fig. 1: Conceptual Relationships

Recommendation System with Multi-Dimensional and Parallel-Case Four-Term Analogy
(Sakaguchi et al. [3] 2011)

- **Objective:** recommend items that both reflect users' preferences and offer valid unexpected elements at the same time.

- Sakaguchi et al. [3] proposed an internet-based recommendation system based on the four-term analogy in this study. This system recommends items that contain users' preferences and valid surprises by mapping the structure from their past preferences.

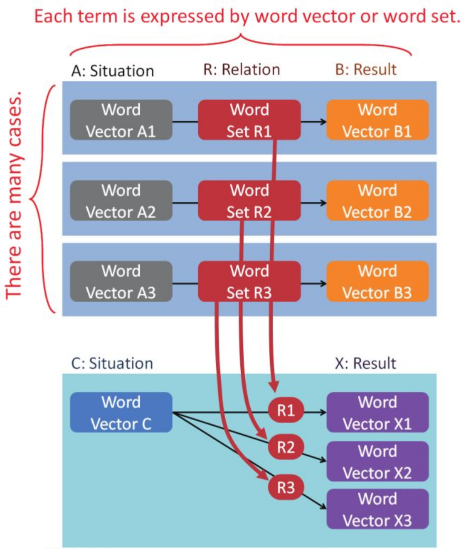


Fig. 2 Multi-dimensional and parallel cases four-term analogy

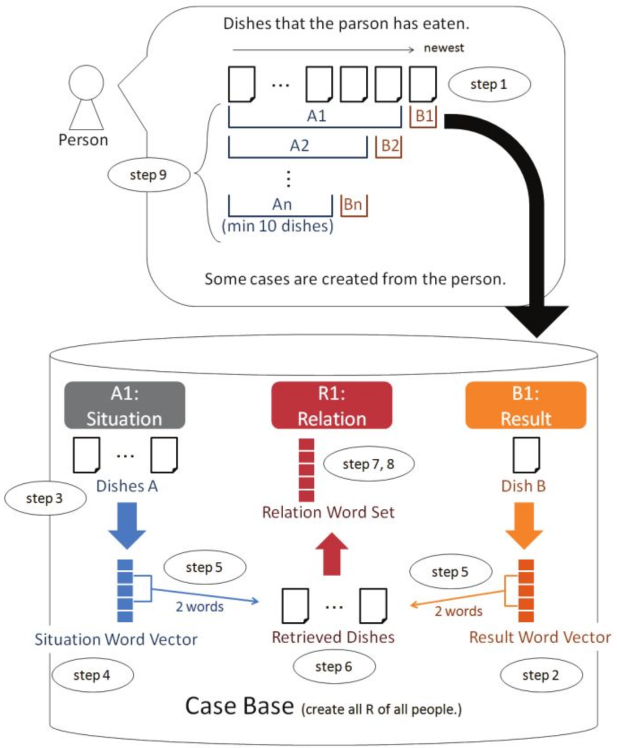


Fig. 4 Flow of case construction

Recipe Recognition with Large Multimodal Food Dataset

(Wang et al. [4] 2015)

- Extended evaluation of **Visual/Textual** features and **Semantic Vector Representation**.

Visual				Textual	Fusion
BoW	Bossanova	Deep	Very Deep	TF-IDF	TF-IDF + Very Deep
23.96%	28.59%	33.91%	40.21%	82.06%	85.10%

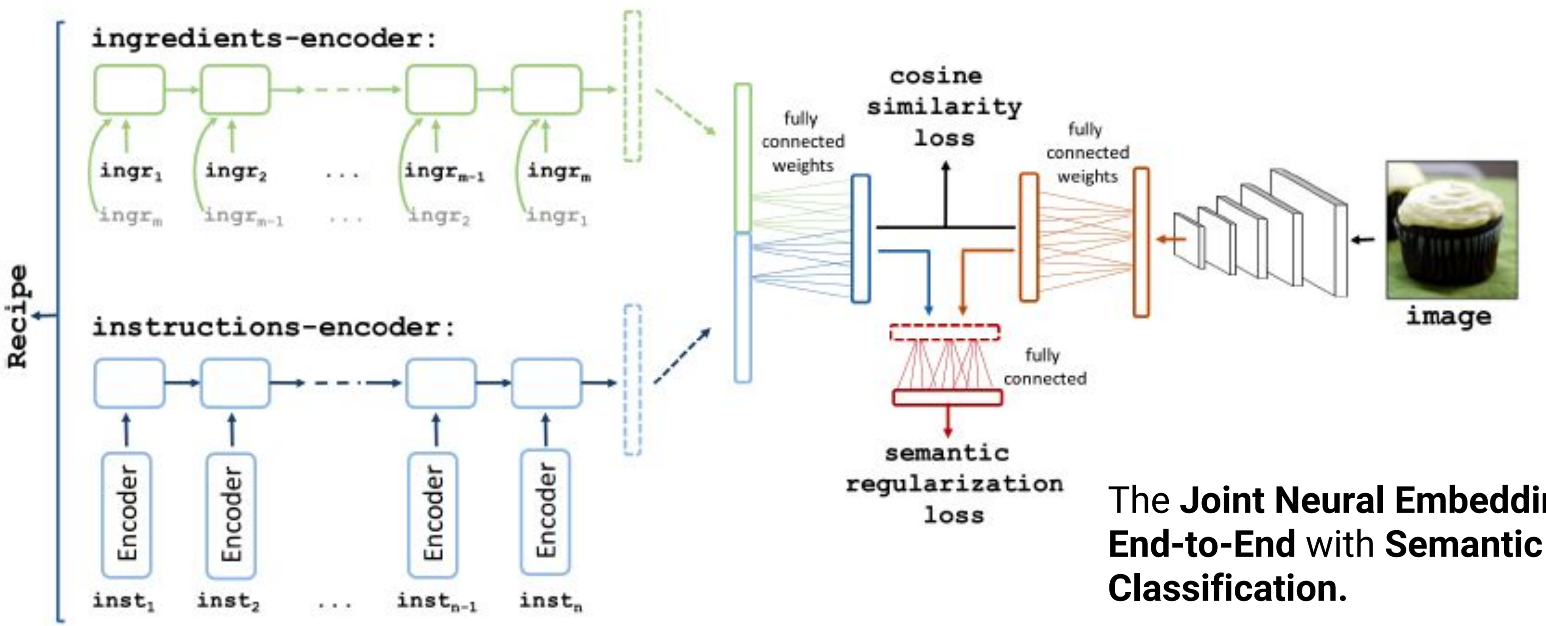
word2vec	TF-IDF+word2vec
67.21%	84.19%

- Semantic Relationships Exploration.
- Recipe Web Search Engine

ravioli	sushi	pho
gnocchi 0.67	nigiri 0.69	souppho 0.68
tortelli 0.58	maki 0.65	vietnames 0.59
cappellacci 0.55	uramaki 0.65	phos 0.57
delallocom 0.52	sashimi 0.64	beefnoodl 0.58
itemtitlea 0.52	norimaki 0.64	bo 0.56

rice	japan	rice japan
calros 0.59	osaka 0.70	**koshihikari 0.64**
basmati 0.59	tokyo 0.62	awabi 0.61
vermicelli 0.58	kyoto 0.62	japanes 0.61
stirfri 0.58	chugoku 0.61	nishiki 0.59
veget 0.58	gunma 0.60	chahan 0.57

Learning Cross-modal Embeddings for Cooking Recipes and Food Images
(Salvador et al. [5] 2017)



Visual Encoding: Resnet-50 without classification layer.

The **Joint Neural Embedding Model** is trained **End-to-End** with **Semantic Regularization by Classification.**

Ingredient Encoding: Bi-directional LSTM.

Instruction Encoding: Skip-Tought [6] followed by LSTM.

Deep Understanding of Cooking Procedure for Cross-modal Recipe Retrieval
(Chen et al. [7] 2018)

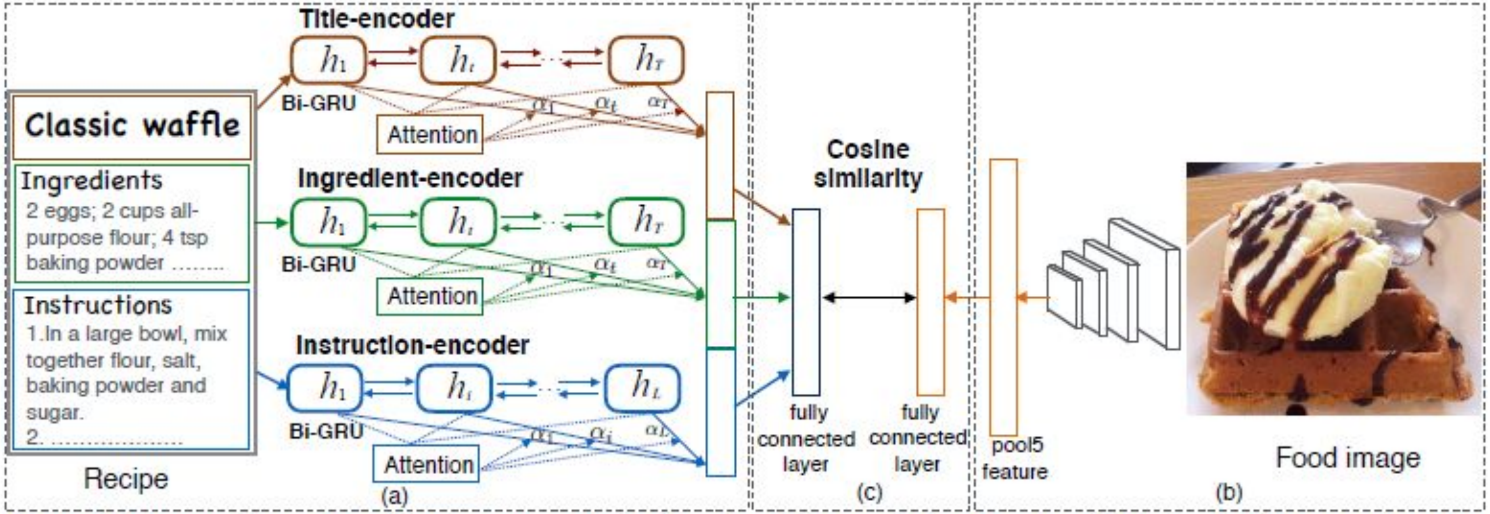


Figure 2: Framework overview: (a) recipe representation learning; (b) image feature learning; (c) joint-embedding space learning.

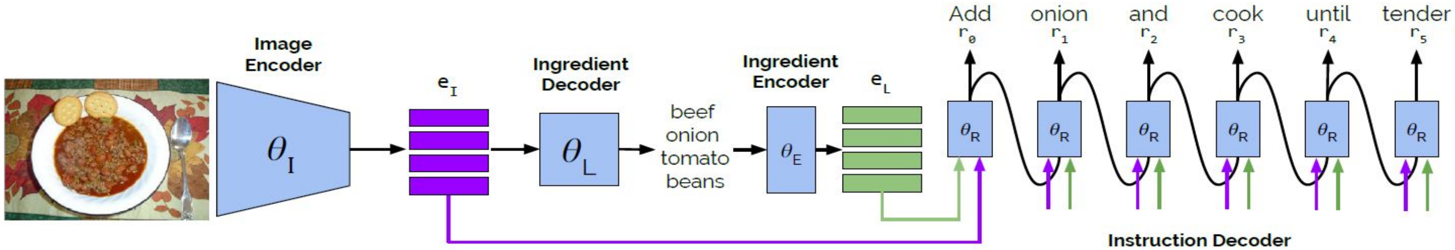
Visual Encoding: Fine-Tuning of Resnet-50.

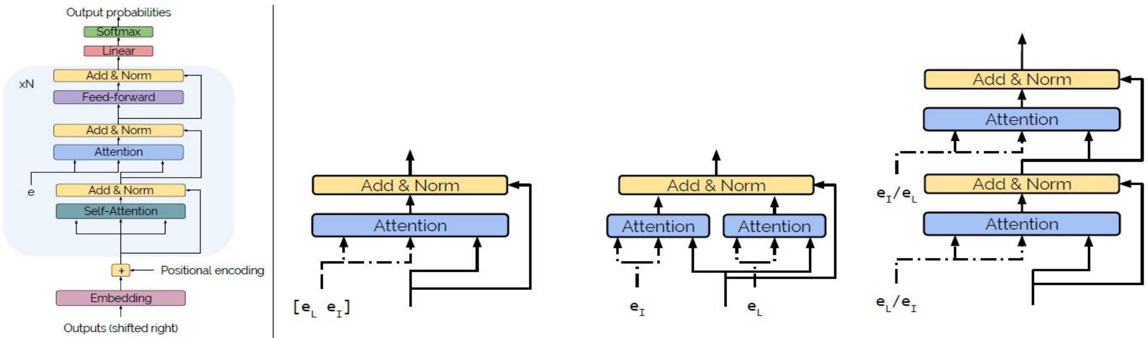
Recipe Encoding: Bi-directional GRU with **Attention Mechanism.**

Attention Mechanism: Models the saliency of words/sentences in recipes and aligns them with the corresponding visual features.

Inverse cooking: Recipe Generation from Food Images

(Salvador et al. [8] 2018)





(a) Transformer model [45] (b) Concatenated (c) Independent (d) Sequential

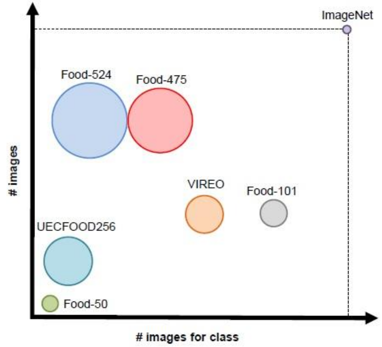
CNN-based Features for Retrieval and Classification of Food Images

(Ciocca et al. [9] 2018)

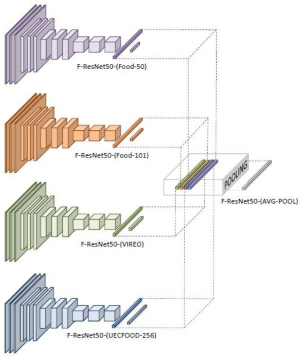
Objective: investigate the use of CNN for the purpose of food recognition and retrieval

Dataset: Food-524 -> Food 475

Network	Top-1 (%)	Top-5 (%)
AlexNet (Food-475)	61.10	84.74
Caffe-Reference (Food-475)	61.43	85.20
GoogleNet (Food-475)	71.75	91.28
VGGNet-16 (Food-475)	73.94	92.28
VGGNet-19 (Food-475)	73.57	93.72
InceptionV3 (Food-475)	74.46	92.95
ResNet-50 (Food-475)	**81.59**	**95.50**
ResNet-50-S (Food-475)	69.45	91.01



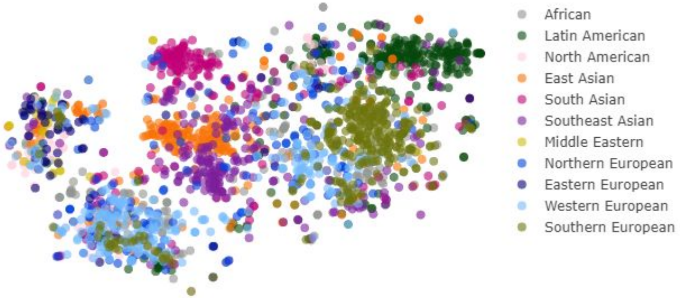
CNN-based Features	UNICT-FD1200 Top-1 (%)	Food-475 Top-1 (%)
F-ResNet-50 (ImageNet)	91.84	40.46
F-ResNet-50 (Food-50)	91.26	37.76
F-ResNet-50 (UECFOOD-256)	94.54	42.17
F-ResNet-50 (Food-101)	95.31	57.95
F-ResNet-50 (VIREO)	94.96	57.92
F-ResNet-50 (AVG-POOL)	95.98	53.69
F-ResNet-50 (Food-524)	**96.56**	67.78
F-ResNet-50 (Food-475)	96.49	**68.01**



Food2Vec (Altosaar [10])

Objective: Build a recommendation system for food & exploring the world's cuisines.

Main idea: embedding of 95,896 recipes



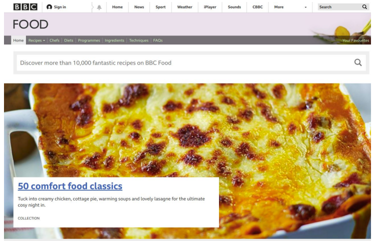
3 tools:

1. Food similarity tool
2. Food analogy tool
3. Recipe recommendation tool

Food A is to food B, as food C is to food D
(king−man)+woman=queen

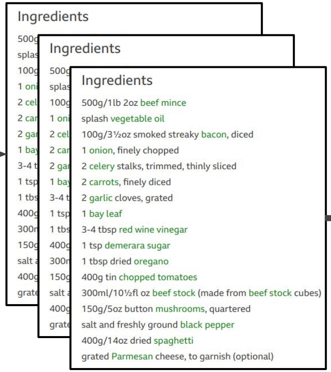
Our Approach

Word Embedding



BBC Food [11]

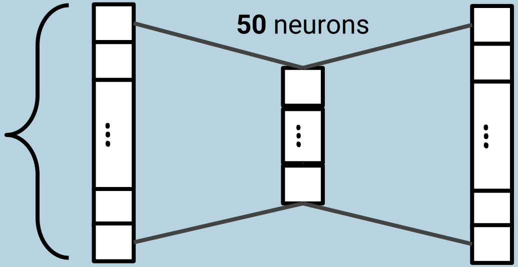
1952 recipes



Ingredients

500g/1lb 2oz beef mince
splash vegetable oil
100g/3½oz smoked streaky bacon, diced
1 onion, finely chopped
2 celery stalks, trimmed, thinly sliced
2 carrots, finely diced
2 garlic cloves, grated
1 bay leaf
3-4 tbsp red wine vinegar
1 tsp demerara sugar
1 tbsp dried oregano
400g tin chopped tomatoes
300ml/10½fl oz beef stock (made from beef stock cubes)
150g/5oz button mushrooms, quartered
salt and freshly ground black pepper
400g/14oz dried spaghetti
grated Parmesan cheese, to garnish (optional)

Auto-Encoder for dimensionality reduction



Binary sparse vectors creations:

$Dish_i$:

Is the **milk** in the ***$dish_i$***?

Representation in the embedded space
(means over dishes for cuisine represented with PCA)

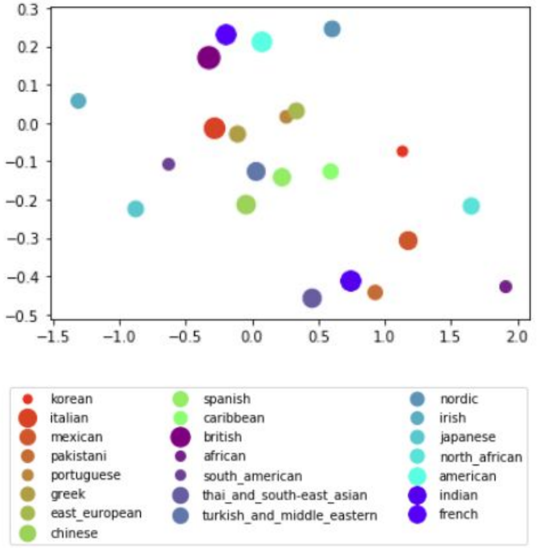
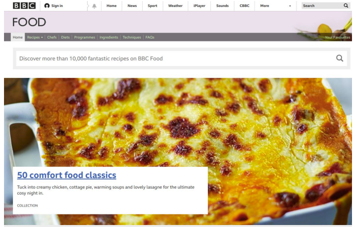


Image Embedding

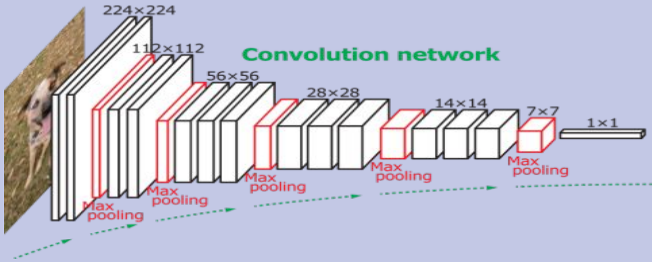
BBC Food [11]

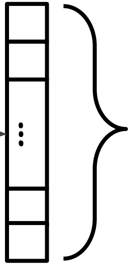


1952 images

MobileNet [12]

(with the output layer removed)

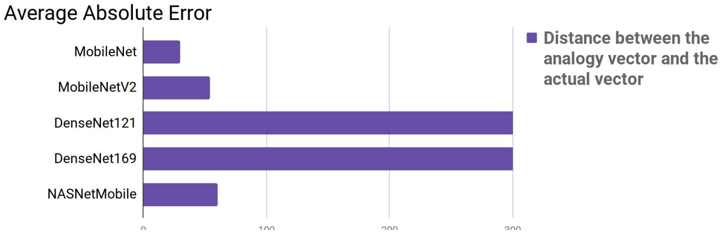




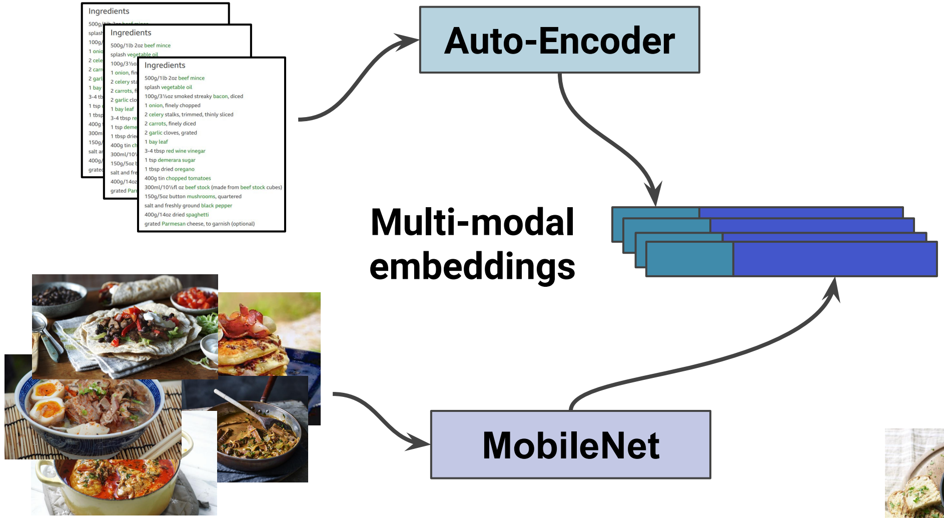
1024 features

Artificial Neural Network selection

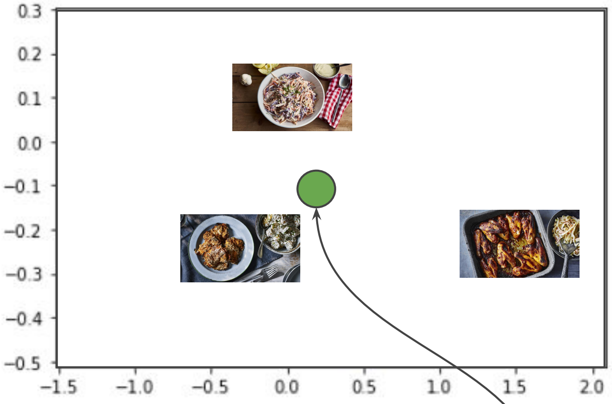
 -  +  ?= 



Analogy Engine



Find the kNN in the embedded space




Selected dish

-


Selected country

+

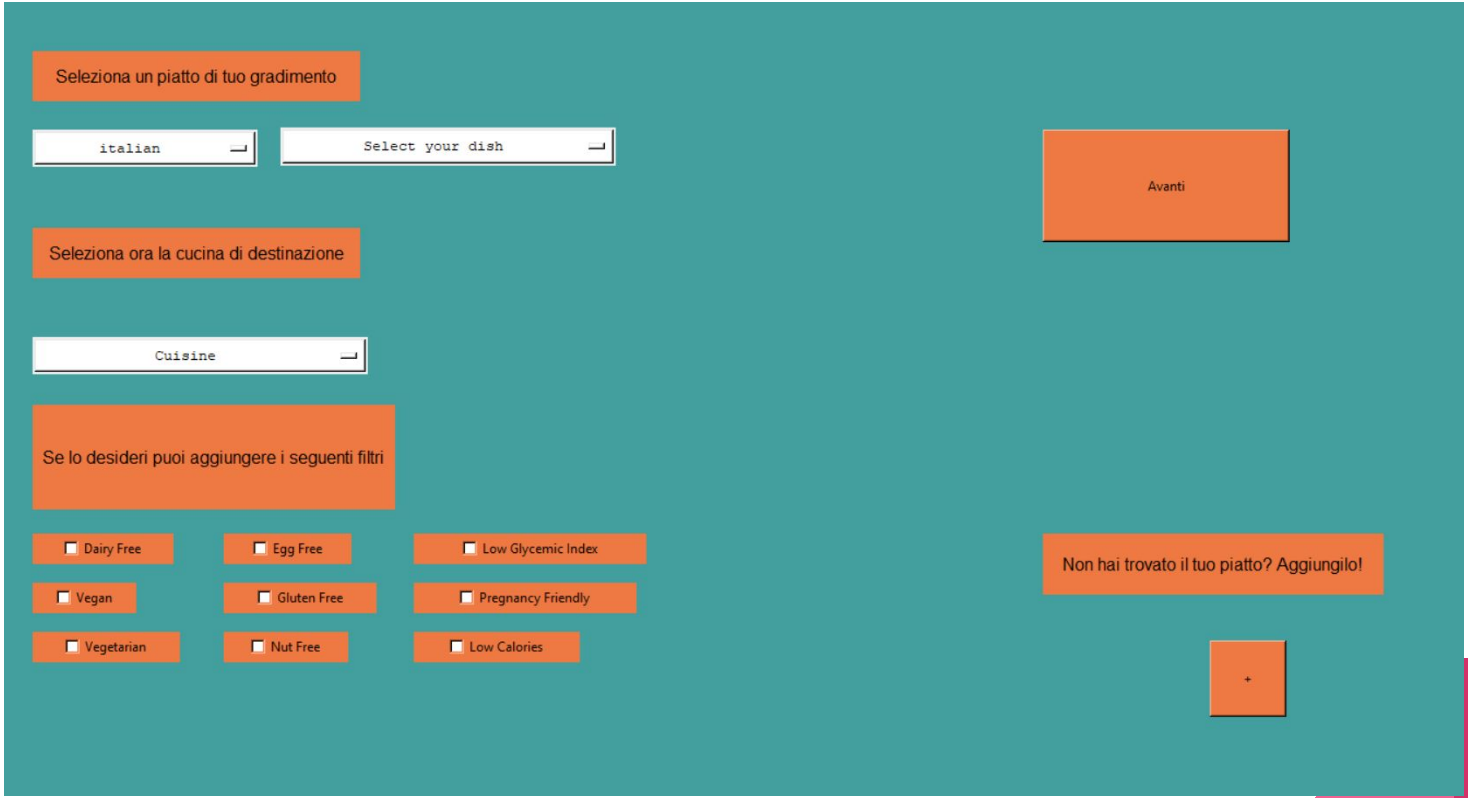

Destination country

=

The user select

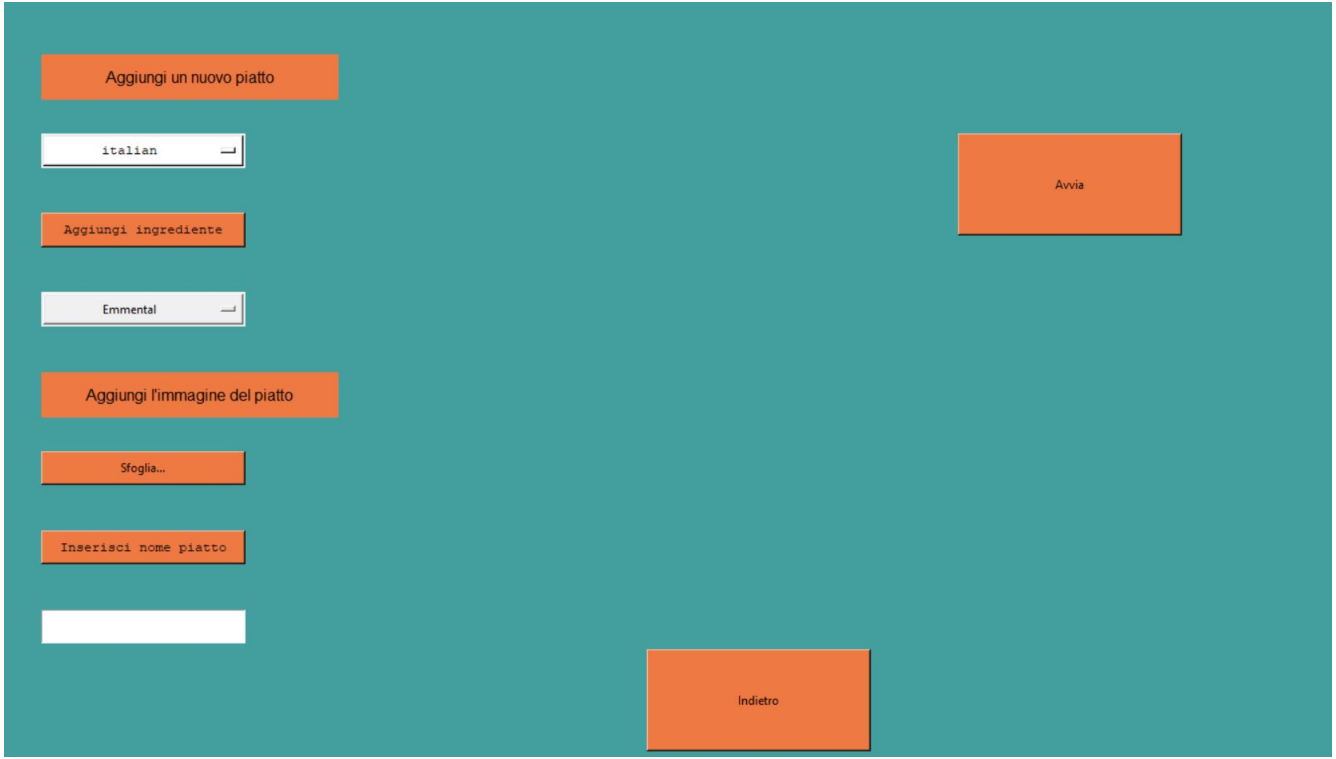
- a dish the he/she likes
- a “destination cuisine”

Graphical User Interface (1)



Graphical User Interface (2)

Graphical User Interface (3)

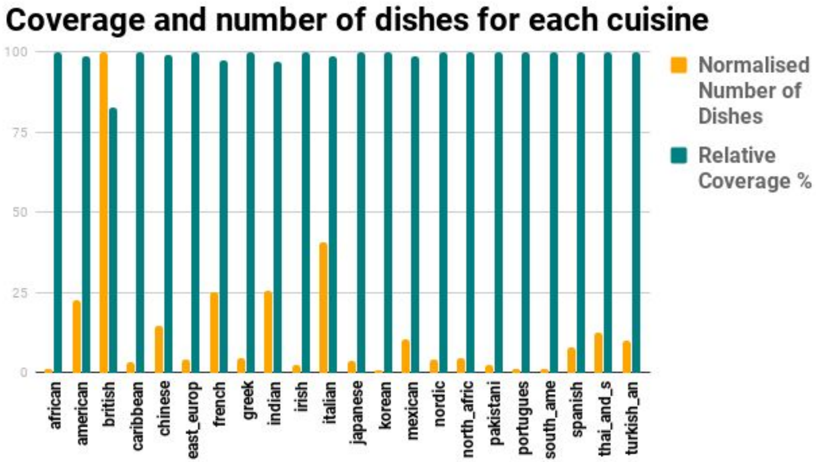


Discussion, conclusion and future development

Discussion

Beef Wllington

Bellini Cocktail



Conclusions and future developments

The results highlight that this approach has interesting capabilities with respect of the coverage metric. However, some problems concerning the recommendation can be found. Furthermore, this be improved in several way, such as:

- Collect more dishes for every cuisine using recipes from local sites (e.g. https://icook.tw/);
- Improve suggestion model by taking in consideration a short description of the cooking procedure;
- Improve the multi-modal embedding procedure;
- Train a CNN end-to-end;
- Improve the GUI aesthetically and functionally;
- Turn our work into an android application.

References

[1]. Hug N., Prade H., Richard G. (2015) Experimenting Analogical Reasoning in Recommendation. In: Esposito F., Pivert O., Hacid MS., Rás Z., Ferilli S. (eds) Foundations of Intelligent Systems. ISMIS 2015. Lecture Notes in Computer Science, vol 9384. Springer, Cham. doi: 10.1007/978-3-319-25252-0_8

[2]. Budán, Paola D., et al. "Recommender system based on argumentation by analogy." *XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014)*. 2014.

[3]. T. Sakaguchi et al., "Recommendation system with multi-dimensional and parallel-case four-term analogy," 2011 IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, 2011, pp. 3137-3143. doi: 10.1109/ICSMC.2011.6084142

[4]. Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015, June). Recipe recognition with large multimodal food dataset. In Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on (pp. 1-6). IEEE.

[5]. Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. *Training*, *720*(619-508), 2.

[6]. Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Skip-thought vectors. In *Advances in neural information processing systems* (pp. 3294-3302).

[7]. Chen, J., Ngo, C., Feng, F., & Chua, T. (2018). Deep Understanding of Cooking Procedure for Cross-modal Recipe Retrieval. *ACM Multimedia*.

[8]. Salvador, A., Drozdzal, M., Giro-i-Nieto, X., & Romero, A. (2018). Inverse Cooking: Recipe Generation from Food Images. *arXiv preprint arXiv:1812.06164*.

[9]. Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding*, *176*, 70-77.

[10]. Jaan Altosaar, Food2Vec, Website (https://jaan.io/food2vec-augmented-cooking-machine-intelligence/), online demo (https://altosaar.github.io/food2vec/#food-analogy-tool) and GitHub repository (https://github.com/altosaar/food2vec)

[11]. Website from which the recipes and the images have been extracted: BBC food https://www.bbc.com/food

[12]. G. Howard Andrew, Zhu Menglong, Chen Bo, Kalenichenko Dmitry, Wang Weijun, Weyand Tobias, Andreetto Marco & Adam Hartwig. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).