AIML/BDA Project Work
A.Y. 2021/22

A multi-modal conversational food recommender system: FoodBot

Castiglia Giovanni and Calò Federica

Outline

1. Scenario and motivation
2. Dataset and features
3. System design and applied methods
4. Experimental evaluation
5. Conclusions

Scenario and motivation

Only a few studies have investigated the use of Conversational Recommender Systems (CRSs) for food recommendation and in particular for encouraging users to make healthier food decisions.

In this project work we have developed a **conversational model** for food recommendation that allows multi-modal and natural user-system interaction.

We compare the impact of two user-system interaction modalities:

- **textual (T)**: communication using only text by displaying the dish names and offering textual explanations of the food recommendations
- **multi-modal (MM):** communication by displaying the name and image of each dish throughout the dialogue, by keeping a textual explanation

Dataset and features

To reach food data, we access the **Allrecipes.com** dataset for educational scopes under EULA agreement and we extracted 2000 recipes and divided them in four food categories:

- Pasta (500 dishes)
- Salad (500 dishes)
- Dessert (500 dishes)
- Snack (500 dishes)

Starting from that, we build our own dataset, composed by the features reported on the right.

Feature names	
id	sodium
category	salt
image_url	tot_grams_weight
name	serving_size
energy_Kcal	proteins_100_grams
energy_Kj	carbohydrates_100_grams
ingredients	fibers_100_grams
servings	sugar_100_grams
proteins	fats_100_grams
carbohydrates	saturates_100_grams
fibers	sodium_100_grams
sugar	salt_100_grams
fats	nutri_score
saturates	fsa_score

System design and applied methods

We developed a system-driven **content-based** food conversational recommender system using *python-telegram-bot* library.

Main steps:

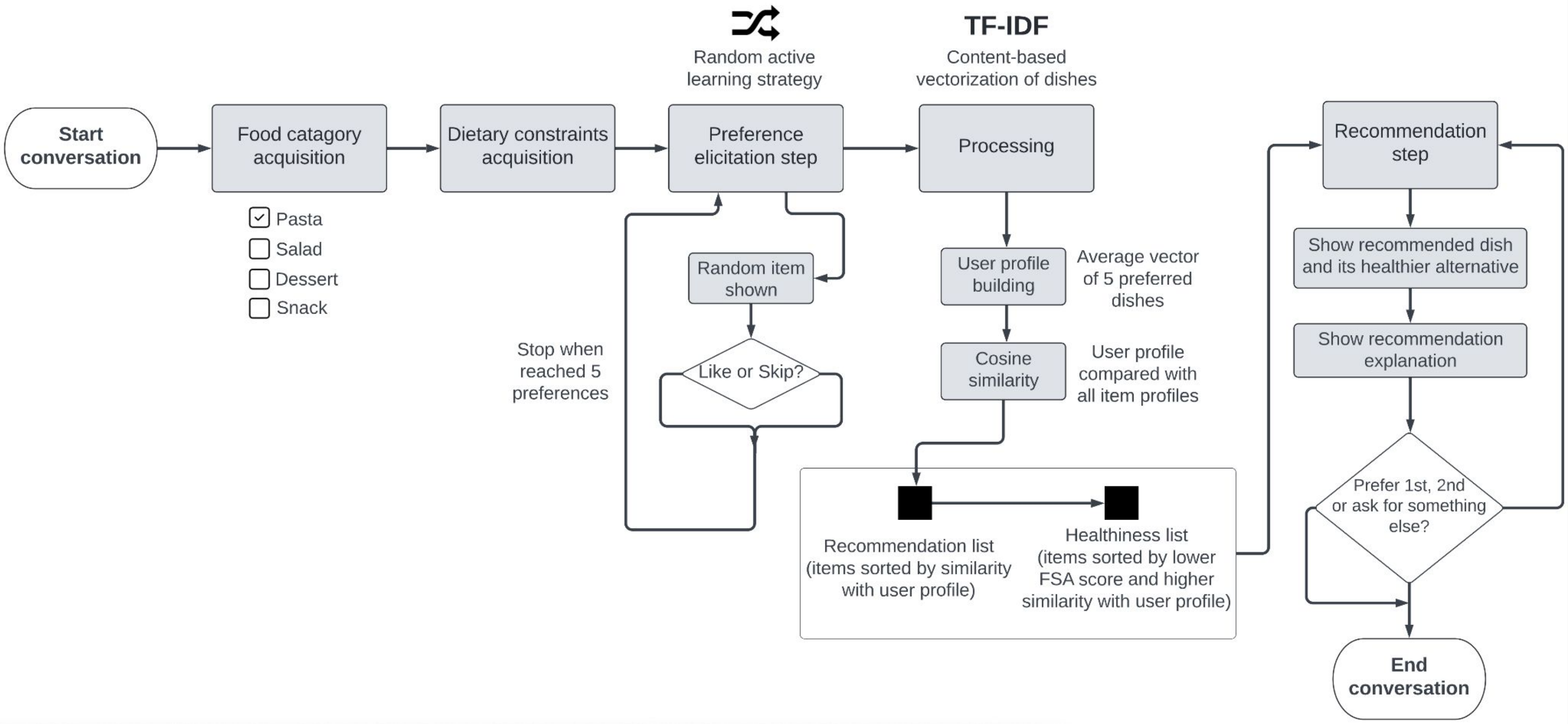
- **Food category acquisition:** choice of food category (Pasta, Salad, Dessert, and Snack)
- **User constraints acquisition:** input of potential dietary constraints (intolerances/diseases or specific ingredients to avoid)
- **Preference elicitation:** submission of preferences for five of the randomly proposed dishes ("Like" and "Skip" buttons for the interaction)
- **Processing:** item representation by TF-IDF. User profile building as mean of the vectors of preferred items. Use of cosine similarity. Extraction of an item from recommendation list (which fits user preferences) and generation of a healthier alternative (rerank based on 60% of healthiness and 40% of similarity with the current recommended dish).
- **Recommendation and explanation:** recommendation of the two dish items and textual explanation which describes why the second dish is healthier than the first one. User can choose one of them or ask for another comparison (iterative process).

System design and applied methods

We sum up here the main methods applied in this project:

- **Random Active Learning strategy:** we use it when proposing dishes during preference elicitation phase. Randomly administered dishes.
- **TF-IDF vectors for item representation:** we represent, for each food category, each dish by a vector of ingredient feature scores using TF-IDF vectorization. In this way we can evaluate the importance of each ingredient in each dish according to all menu items.
- **Cosine similarity for item comparisons:** we compare item vectors with user profile vector (same size and features) by performing the cosine similarity score. This help us ranking dishes similar to user profile (preferences)
- **Reranking based on features (i.e. similarity score and fsa_score)**

System design and applied methods



Experimental evaluation

We run a within-subject user study with N=30 participants (family-and-friends users).

Users were asked to interact with the two versions of our chatbot (which differ in interaction modality, T or MM). Then, they were asked to fill a simple post-task questionnaire on Google Forms which evaluated the user's experience through choice satisfaction (which modality is <u>more satisfying</u>), effectiveness (which modality is <u>easier to use</u>) and system efficiency (which modality is <u>faster to use</u>), using 5-point Likert scales.

We obtained a 100% agreement of users on the preferred modality: MM (multi-modal)
The main reasons of their choice are <u>clarity</u> of text + images interaction wrt pure textual interaction and <u>good quality</u> of images

Experimental evaluation

As regards to Satisfaction, Effectiveness and Efficiency measures, we obtained that the majority of users prefer MM modality.

We also performed a T-test for investigating if the interaction modality (T and MM groups) impacts these variables:

- conversation duration
- preference elicitation step duration
- recommendation step duration
- number of interactions
- has user performed a healthier choice
- FSA score of selected dish
- number of skips during preference elicitation step

We do not obtained p-values that allow to reject the null hypothesis (<0.05). The variables in red have slightly not reached this goal. The main reason of this failure could be considered the limited number of users.

Experimental evaluation

Other insights from our analysis are reported below:

- with both interaction modalities, the majority of user performed healthier choices at the end of the conversation (by choosing the second healthier alternative).
- conversation duration (in particular recommendation step duration) is lower with MM modality: this tells us that images help user deciding faster about their tastes.
- the preference elicitation duration is higher with MM modality and this is caused by a higher number of skips. This could infer that users with images are allowed to define more accurate preferences wrt textual interaction.
- probably, with a higher number of participants we could have stated that <u>conversation duration</u> and <u>number of interactions</u> variables could be influenced by the interaction modalities.

Conclusions

In conclusion, we developed a **multi-modal** conversational approach for healthy food recommendation. We were able to test the functionalities of our system and investigate the impact of the two versions (T and MM interaction mods.) on user satisfaction, effectiveness and efficiency and on system measurements (timings, etc.).

Unfortunately we could not infer that modalities have a direct impact on the variables due to T-test incomes. This is probably caused by the limited number of the user study participants.

We applied statistical analysis and techniques learned from the AIML/BDA course in this project work and we equally distributed the work tasks.

GitHub repo: https://github.com/giocast/ML-Project-FoodBot